

EIN ALLGEMEINES MODELL FÜR DIE SKALIERUNG VON
MEINUNGEN UND EINSTELLUNGEN

I n a u g u r a l - D i s s e r t a t i o n

zur

Erlangung des akademischen Grades eines

Doktors des Wirtschaftswissenschaften

durch die

Rechts- und Staatswissenschaftliche Fakultät

der Universität Münster

vorgelegt von

Dieter P. Wehowski

aus

Breslau

1961



Digitized by the Internet Archive
in 2026

EIN ALLGEMEINES MODELL FÜR DIE SKALIERUNG VON
MEINUNGEN UND EINSTELLUNGEN

I n a u g u r a l - D i s s e r t a t i o n
zur

Erlangung des akademischen Grades eines
Doktors der Wirtschaftswissenschaften
durch die

Rechts- und Staatswissenschaftliche Fakultät
der Universität Münster .

vorgelegt von

Dieter P. W e h o w s k i

aus

Breslau

2312 / 167

1. Berichterstatter: Dozent Dr.H.König
 2. Berichterstatter: Professor Dr.H.Schelsky
- Dekan: Professor Dr.P.Deutsch
- Tag der mündlichen Prüfung: 18. Januar 1961

Die vorliegende Arbeit wurde von Herrn Professor Dr.H.Kellerer angeregt. Ihm, wie auch Herrn Professor Dr.W.G. Hoffmann und Herrn Dozent Dr.E.König möchte ich für Ihr Interesse und wertvolle Hinweise danken.

"That ... only illustrates the
hard lot of the economist,
masquerading as a man of science
in a universe not of cells and
atoms but of passions and
volitions."

Sir Dennis Robertson
"The Snake and the Worm"

Inhaltsverzeichnis

	Seite
Von der Arbeit	v
Einleitung	1
<u>1. Kapitel</u>	
Problemstellung und Überblick	1
<u>2. Kapitel</u>	
Die Konstruktion einer topologischen Skala	21
<u>3. Kapitel</u>	
Die Konstruktion einer metrischen Skala	7
<u>4. Kapitel</u>	
Ergebnisse und Ausblick	14
Literaturverzeichnis	143

a. In den letzten Jahren wurden unter dem Namen "Scaling" eine Reihe von statistischen Methoden entwickelt, deren praktische Bedeutung sehr groß sein kann. Jedoch sind sie bis jetzt noch wenig bekannt. Es handelt sich dabei, allgemein gesprochen, um Methoden für die Messung und damit Quantifizierung "qualitativer, psychischer Eigenschaften" eines oder einer Gruppe von Menschen. Diese "Eigenschaften" können Einstellungen zu bestimmten Produkten, Meinungen, Fähigkeiten oder auch zB. der "subjektive Nutzen" sein. Der Raum, innerhalb dessen Scaling oder, wie wir sagen wollen, die Skalierungstechnik angewendet werden kann, reicht also von den praktischen Bedürfnissen der Markt- und Meinungsforschung bis zu der Verifikation ökonomischer Hypothesen (zB. aus der Nutzentheorie).

Skalieren heißt allgemein, die Mengen einer Eigenschaft auf die Menge der reellen Zahlen abbilden. Diese Abbildung muss so erfolgen, daß die Beziehungen zwischen den Maßzahlen die Beziehungen zwischen den entsprechenden, gegebenen Mengen der Eigenschaft so weitgehend als möglich widerspiegeln. Je nachdem, welche Struktur die Eigenschaft hat und welche Methode der Skalierung man anwendet, werden entweder aus der Skala nur Rückschlüsse auf die Ordnung der korrespondierenden Mengen möglich sein, oder aber es wird sich eine metrische Skala konstruieren lassen; das heißt, die Zahlen sind dann derart zugeordnet, daß die Maßzahl einer Summe von Mengen gleich ist der Summe der Maßzahlen.¹

¹Pfanzagl, J.: Die axiomatischen Grundlagen einer allgemeinen Theorie des Messens, Würzburg, 1959.

Plan der Arbeit

Einführung

1. Kapitel: Problemstellung und Überblick

- A. Definition des "Messens"
 - 1. Begriff des Messens
 - 2. Objekt der Messung
- B. Methoden der Messung
 - 1. Grundgedanke
 - 2. Abriss des Modells
 - 3. Problemstellung
 - 4. Zusammenfassung
- C. Überblick über den Aufbau der Arbeit

2. Kapitel: Die Konstruktion einer topologischen Skala.

- A. Die Auswahl der Items
 - 1. Begriff und Aufbau des "Item"
 - 2. Mögliche Auswahlkriterien
 - a. Kriterien für die Formulierung der Items
 - b. Statistische Auswahlkriterien
 - 3. Auswahlmethoden
 - a. Vorbemerkung
 - b. Grundgedanke der Methoden
 - aa. Sinn der Berechnung von α_1 und α_2
 - bb. Sinn der Berechnung von α_j
 - cc. Die Größe der Beurteilergruppe
 - c. Aufbau der Auswahlmodelle (I)
 - aa. Die Aufgabe der Beurteiler
 - bb. Das erhaltene Datenmaterial
 - cc. Die Berechnung von α_j und α_2
 - dd. Interpretation der Ergebnisse
 - d. Aufbau der Auswahlmodelle (II)
 - aa. Aufgabe der Beurteiler
 - bb. Das erhaltene Datenmaterial
 - cc. Berechnung von α_j
 - dd. Interpretation der Ergebnisse
 - 4. Zusammenfassung
 - a. Die Ergebnisse der Itemauswahl
 - b. Hinweise auf die Auswahl der Befragten
- B. Die Konstruktion der Skala
 - Vorbemerkung
 - a. Grundgedanke
 - b. Die verwendeten Methoden
 - I. Kategorische Methoden
 - 1. Grundgedanke
 - 2. Auswertungsmethoden
 - a) Das Skalogramm-Brett
 - b) Die Cornell-Technik
 - c) Hinweis auf die Lösung des Problems nach der Methode der kleinsten Quadrate

3. Der Test auf Eindimensionalität
 - a. Reproduzierbarkeit (I)
 - aa. Berechnung des Maßes
 - bb. Interpretation der Ergebnisse
 - b. Reproduzierbarkeit (II)
 - aa. Grundgedanke
 - bb. Berechnung des Maßes
 - c. Zusammenfassung
 4. Die Bestimmung des O-Punktes
 - a. Grundgedanke
 - b. Auswertung des Datenmaterials
 - c. Beispiel für Verzerrungsfreiheit des O-Punktes
 5. Zusammenfassung
- II. Komparative Modelle
1. Grundgedanke
 - a. "Objektive" Skala versus individuelle Präferenzskalen
 - b. Aufgabe
 2. Gegeben: totale Präferenzskalen
 - a. Bestimmung der Rangordnung der Befragten
 - b. Die erzielte teil-metrische Information
 - c. Beispiel
 3. Gegeben: partielle Präferenzskalen
 - a. Aufgabe der Befragten
 - b. Die erzielte Information
 - aa. Topologische Informationen
 - bb. Teil-metrische Informationen
- C. Zusammenfassung
1. Die erzielten Ergebnisse
 - a. Der "Maßstab"
 - b. Die Messung der Verteilung
 - aa. Die topologische Messung
 - bb. Der O-Punkt
 - c. Der Test auf Eindimensionalität
 2. Weitere Aufgaben
 - a. Die Mängel einer topologischen Messung
 - b. Die Möglichkeiten einer metrischen Skalierung
3. Kapitel: Die Konstruktion einer metrischen Skala
- A. Das theoretische Modell für die Bestimmung des Skalenwerts der Items oder Stimuli
1. Der Grundgedanke des Modells
 - a. Die Aufgabe der Beurteiler
 - b. Die Voraussetzungen des Modells
 - c. Das Modell
 2. Vereinfachende Annahmen
 - a. Konstante Standardabweichungen der Verteilungen der Reaktionsprozesse
 - b. Geringe Differenzen in den Standardabweichungen der Reaktionsprozesse
 3. Zusammenfassung

- B. Die praktische Bestimmung des Skalenwerts der Items oder Stimuli
 - 1. Das empirische Material
 - a. Die Sammlung der Daten
 - aa. Die Aufgabe der Beurteiler
 - bb. Die Zahl der Beurteiler
 - cc. Darbietung der Stimuli
 - dd. Zum Begriff des "Stimulus"
 - b. Die Grundmatrizes
 - aa. Die H-Matrix
 - bb. Die P-Matrix
 - cc. Die M-Matrix
 - 2. Die Bestimmung des Skalenwerts (I)
 - a. Die praktische Bestimmung
 - b. Beweis für die Berechtigung des Vorgehens
 - c. Vorgehen bei unvollständiger Matrix
 - d. Zusammenfassung
 - 3. Bestimmung des Skalenwertes (II)
 - a. Die Ableitung der Formel
 - b. Vorgehen bei unvollständiger Matrix
 - c. Zusammenfassung
 - 4. Praktisches Beispiel
 - a. Das Datenmaterial
 - aa. Die Items
 - bb. Die Antwortmatrizes
 - b. Berechnung des Skalenwertes für den Fall (I)
 - c. Berechnung des Skalenwertes für den Fall (II)
 - d. Berechnung des Skalenwertes bei unvollständiger Matrix
- C. Test auf die Richtigkeit der zugrundeliegenden Hypothesen
 - 1. Vorbemerkung
 - 2. Ein Additivitätstest
 - a. Grundgedanke
 - b. Ein einfaches Maß für die Abweichungen
 - c. Anwendung des χ^2 Verfahrens
 - aa. Die Hypothesen
 - bb. Die arcsin Transformation
 - d. Beispiel
 - 3. Ein Konsistenztest
 - a. Grundgedanke
 - aa. Die Triaden
 - bb. Interpretation
 - b. Die Berechnung der Maßzahl
 - aa. Eine einfache Maßzahl
 - bb. Anwendung des χ^2 Verfahrens
 - c. Beispiel
 - 4. Zusammenfassung
 - a. Ergebnisse des Additivitätstests
 - b. Ergebnisse des Konsistenztests
- D. Die Messung der Einstellung
 - 1. Die Skalierung der Befragten

- a. Grundgedanke der Methode
- b. Drei Möglichkeiten der Skalierung
- c. Ergebnisse
- 2. Die Bestimmung des 0-Punktes
 - a. Durch eine Intensitätsanalyse
 - b. Durch die Methode der summativen Beurteilung
- 3. Zusammenfassung

4. Kapitel: Ergebnisse und Ausblick

- 1. Ergebnisse der Untersuchung
 - a. Das Problem
 - b. Die Lösung
- 2. Ausblick auf weitere Forschungsmöglichkeiten
 - a. Allgemeine Möglichkeiten
 - b. Zum Problem mehrdimensionaler Modelle

Literaturverzeichnis

* * * * *

Die Methoden, derer man sich bei der Konstruktion solcher Skalen bedienen kann, sind die der Beurteilung und Befragung. Sie basieren auf der Annahme, daß es möglich ist, aus einer großen Zahl in bestimmter Form abgegebener, subjektiver Bewertungen zu objektiven Aussagen und Meßresultaten zu gelangen. Viele der Fragestellungen der Skalierungstechnik sind so im Grunde identisch mit alten Problemen der Psychophysik und Psychometrie.

In der vorliegenden Arbeit wird versucht ein allgemeines Modell für die Skalierung von eindimensionalen Eigenschaften zu konstruieren. Das Modell wird allgemein sein in dem Sinne, daß es sowohl eine topologische (rangmäßige) wie metrische Skalierung erlaubt. Mit eindimensional bezeichnet man Eigenschaften dann, wenn die Anordnung von Objekten nach dieser Eigenschaft nur in einer einzigen Weise möglich (invariant) ist.

b. Das im Folgenden entwickelte Skalierungsmodell ist ganz auf die praktischen Belange der Marktforschung wie der Meinungsforschung zugeschnitten. Hier dürfte sein wichtigstes Anwendungsgebiet liegen; zum Beispiel bei der Messung von Kaufmotiven, des Werbeerfolgs usw.. Ehe wir jedoch zu seiner Darstellung übergehen, sei noch folgender Hinweis erlaubt.

Es erscheint intuitiv plausibel mit der hier vorgeschlagenen Methode an das alte Problem der Nutzenmessung heranzugehen. Die im 2.Kapitel (B,II) dargestellte Methode der Auflösung subjektiver Präferenzskalen und noch mehr eine

Technik der Bewertung von Alternativen, ähnlich der in Kapitel 3,B entwickelten Methode sind dazu geeignet.

Im zweiten Fall suchen wir nach einer numerischen Skala mit folgender Eigenschaft: die Wahrscheinlichkeit dafür, daß eine Alternative der anderen vorgezogen wird, hängt nur von der Differenz der Skalenwerte der betreffenden Alternativen ab. D.h., wenn s die Skala bezeichnet und $P(i,j)$ die Wahrscheinlichkeit, daß i j vorgezogen wird, dann ist für $P(i,j) \neq 0,1$

$$P(i,j) = P'[s(i) - s(j)]$$

Für auf eindimensionalen Kontinua gegebene Alternativen ist durch diese Bedingung eine metrische Skala bestimmt. Nicht jedoch, wenn wir es mit diskreten Alternativen zu tun haben, wie es bei Präferenzexperimenten in der Nutzenmessung der Fall ist. Um hier ebenfalls zu einer (bis auf lineare Transformationen) eindeutigen Skala zu gelangen, müssen die Wahrscheinlichkeiten P gewissen restriktiven Bedingungen genügen und die Experimentanordnung eine reiche mathematische Struktur des Ausgangsmaterials gewährleisten. Wir werden sehen, daß wir unter diesen Voraussetzungen das Problem der Nutzenmessung auf das hier vorgetragene reduzieren können.

Im folgenden werden kurz die axiomatischen Grundlagen eines Modells skizziert, das die Nutzenmessung mithilfe der hier vorgeschlagenen Skalierungstechnik ermöglicht. Wir bauen dabei hauptsächlich auf den Gedanken von Neumann-Morgenstern,

R.D.Luce und J.Pfanzagl auf.¹

Es sei A eine (finite oder infinite) Menge reiner Alternativen und E eine Menge von Zufallselementen. Daraus lässt sich die Menge S von kombinierten Situationen bilden. Für $a, b \in A$ und $\omega \in E$, würde die kombinierte Situation $a \omega b$ besagen: a wenn das Zufallsereignis ω eintritt und b wenn es nicht eintritt.

Wir nehmen nun an, daß eine objektive Wahrscheinlichkeit $w(a, b)$ für $a, b \in A$ dafür besteht, daß eine bestimmte Person a b vorzieht. Um die starken Transitivitätsbedingungen der Neumann-Morgensternschen Theorie zu umgehen, fordern wir mit Luce, daß - quasi induziert - $a \succeq b$ dann, wenn für jedes $c \in S$

$$A I \quad w(a, c) \geq w(b, c) \text{ und } w(c, b) \geq w(c, a)$$

Voraussetzung ist natürlich, da Vergleiche immer möglich sind. Das heißt für jedes $a, b \in S$ entweder

$$A II \quad a \succeq b \quad \text{oder} \quad b \succeq a$$

¹Das hier verwendete Nutzenkonzept entspricht nicht der traditionellen Theorie des Nutzens in risikofreien Situationen, die, wie Savage schreibt "has been completely discredited in the eyes of almost all economists" (L.J.Savage: Foundations of Statistics, New York, 1954, S.96). Ein übersichtlicher Abriss dieses Teils der Nutzentheorie findet sich bei Stigler (G.J.Stigler: The Development of Utility Theory, Journ.Pol.Ec., 58, 1950, 307-327, 377-396.), sie wird kritisiert von Pareto (V.Pareto: Manuel d'économie politique, 2.Aufl., Paris, 1927) Savage (L.J.Savage, a.a.O., 1954), Neumann-Morgenstern (J.v.Neumann und O.Morgenstern: Theory of Games and Economic Behavior, 2nd ed., Princeton, 1947), Bohnert (H.C.Bohnert: The Logical Structure of the Utility Concept, in R.M.Thrall a.oth., eds., Decision Processes, New York, 1957) und anderen. Unser Nutzenkonzept entspricht dem der modernen Entscheidungstheorie. Hier ist der Nutzen eine Funktion, die Präferenzbeziehungen unter Unsicherheitsbedingungen arithmetisiert. (F.P.Ramsey: The Foundations of Mathematics, New York, 1931; J.v.Neumann und O.Morgenstern, a.a.O., 1947; R.D.Luce: A Probabilistic Theory of Utility, Econometrica, 26, 1958, 193-214; J.Pfanzagl, a.a.O., 1959 und andere).

Weiterhin müssen wir fordern, daß sozusagen kein Nutzen des Spiels selbst besteht. Das wird durch folgende Relation ausgedrückt: Für $a, b, c \in A$ und $\alpha, \beta \in E$

$$A \text{ III} \quad (a \alpha b) \beta c \sim a (\alpha \cap \beta) (b \beta c)$$

Das heißt, auf beiden Seiten ist a das Ergebnis wenn $\alpha \cap \beta$
 b wenn $\alpha \cap \beta$ und c wenn $\bar{\beta}$ eintritt. Da $a \alpha a = a$
 impliziert A III

$$(a \alpha b) \beta b \sim a (\alpha \cap \beta) b$$

was bedeutet, daß gewisse zweistufige Spiele einstufigen äquivalent sind.

Nehmen wir nun an, dem Befragten würden zwei Situationen $a \alpha b$ und $a \beta b$ zur Auswahl vorgelegt. Seine Entscheidung wird davon abhängen, welche Alternative er bevorzugt und welches der Zufallsereignisse er für wahrscheinlicher hält. Mithin sind vier Kombinationen möglich, von denen zwei zur Bevorzugung von $a \alpha b$ über $a \beta b$ führen müssten: wenn $a \alpha b$ vorgezogen wird und α für wahrscheinlicher gehalten als β - und wenn $b \alpha a$ vorgezogen und β für wahrscheinlicher als α gehalten wird. Wenn wir für die Wahrscheinlichkeit, daß α für möglicher gehalten wird als β $q(\alpha, \beta)$ schreiben, so ist die Wahrscheinlichkeit dafür, daß $a \alpha b$ $a \beta b$ vorgezogen wird

$$A \text{ IV} \quad w(a \alpha b, a \beta b) = w(a, b) q(\alpha, \beta) + w(b, a) q(\beta, \alpha)$$

Dieses Axiom enthält mehrere Implikationen. Vor allem aber, daß die zwei Diskriminanzprozesse voneinander unabhängig sind. Das heißt also, daß der Befragte die Alternativen a und b

für unabhängig von α und β ansieht.

Die nun folgenden Axiome A V bis A VII sind relativ harmlos. Es sei

$$A \underline{V} \quad w(a, b) \geq 0 \quad \text{und} \quad w(a, b) + w(b, a) = 1$$

$$q(\alpha, \beta) \geq 0 \quad \text{und} \quad q(\alpha, \beta) + q(\beta, \alpha) = 1$$

Das heißt, w und q sind Wahrscheinlichkeiten, die zwischen 0 und 1 liegen, und der Befragte ist gezwungen, eine Wahl zu treffen. Auch diese Bedingung entspricht der später angewendeten Technik.

Aus A IV und A V ergibt sich übrigens

$$q(\alpha, \beta) = \frac{w(a\alpha b, a\beta b) + w(a, b) - 1}{2w(a, b) - 1}$$

Diese Formel ermöglicht die Berechnung der $q(\alpha, \beta)$ aus den Testdaten. Wir nehmen an, es existiere in S wenigstens eine Funktion u , die Nutzenfunktion, und wenigstens eine Funktion ϕ in E , die Funktion der subjektiven Wahrscheinlichkeiten, für die

$$A \underline{VI} \quad u(a) \geq u(b) \quad \text{genau dann, wenn} \quad a \succeq b \quad \text{für} \quad a, b \in S$$

$$\phi(\alpha) \geq \phi(\beta) \quad \text{genau dann, wenn} \quad \alpha \succeq \beta \quad \text{für} \quad \alpha, \beta \in E$$

Ein vertrautes Konzept der Nutzentheorie.

Für unsere spätere Beweisführung ist es nun noch wichtig, A IV in bestimmter Weise zu erweitern

$$A \underline{VII}, 1 \quad w[(a \gamma b) \alpha b, (a \gamma b) \beta b] = w(a \gamma b, b) q(\alpha, \beta) + w(b, a \gamma b) q(\beta, \alpha)$$

Das heißt, nicht nur a und b sind unabhängig von α und β , sondern auch α und β von γ . Wir können dann schreiben

$$A_{VII,2} \quad q(\alpha \cap \beta, \gamma) = q(\alpha, \beta) q(\gamma, 0) + q(\beta, \alpha) q(0, \gamma)$$

Ferner haben die Funktionen der subjektiven Wahrscheinlichkeiten die Eigenschaft, daß für alle $x, y, z \in [0, 1]$

Ereignisse α, β, γ in E bestehen, so daß

$$A_{VIII} \quad \phi(\alpha) = x, \quad \phi(\beta) = y, \quad \phi(\gamma) = z$$

und α und β unabhängig von γ sind. Das heißt, ϕ ist dicht.

Die zwei subjektiven Skalen erfüllen die Nutzenhypothese in dem Sinn, daß für $a, b \in A$ und $\alpha \in E$

$$A_{IX} \quad u(\alpha \alpha b) = \phi(\alpha) u(a) + \phi(\bar{\alpha}) u(b)$$

Ebenfalls eine vertraute Gleichung der Nutzentheorie.

Anhand dieser Axiome nun läßt sich zeigen, daß

$$\begin{aligned} q(\alpha, \beta) &= q'[\phi(\alpha) - \phi(\beta)] \\ u(a, b) &= u'[\phi(a) - \phi(b)] \quad ^1) \end{aligned}$$

¹Für den Beweis siehe vor allem Luce, a.a.O., S.209. Auf Grund von A VI bis A VIII schreiben wir A VII,2

$q(xz, yz) = q(xy) q(z, 0) + q(y, x) q(0, z)$
 q' sei auf $[-1, 1]$ wie folgt definiert

$$q'(x) = \begin{cases} q(x, 0) & \text{wenn } x \in [0, 1] \\ q(0, -x) & \text{wenn } x \in [-1, 0] \end{cases}$$

Angenommen $x \geq y$ so ist $y/x \leq 1$, dann ist

Genauso $q(x, y) = q(x, y/x, \frac{y}{x}) = q(1, \frac{y}{x}) q(x, 0) + q(\frac{y}{x}, 1) q(0, x)$

Da $q'(x-y) = q(x-y, 0) = q[(1-\frac{y}{x})x, 0] = q(1-\frac{y}{x}, 0) q(x, 0) + q(0, 1-\frac{y}{x}) q(0, x)$

$\phi(\alpha) + \phi(\bar{\alpha}) = 1$ und $q(e, 0) = 1$ und $q(0, e) = 0$

(e sei das universelle Ereignis der Booleschen Algebra E)
 ergibt sich

Das heißt, die Wahrscheinlichkeit dafür, daß $a \succ b$ und $a \succ b$ vorgezogen wird, ist eine Funktion der Differenz der respektiven Skalenwerte. Damit ist das Problem der Nutzenmessung auf das in Kapitel 3 geschilderte Problem der Konstruktion einer metrischen Skala reduziert. Nach noch einigen weiteren Annahmen kann auch die Form der Funktion bis auf einen Parameter festgelegt werden. Die zwei subjektiven Skalen können dann über die empirischen Schätzwerte für die Wahrscheinlichkeiten w bestimmt werden.

c. Soviel zum Problem der Nutzenmessung. Kehren wir zu unserem eigentlichen Ziel, der Konstruktion eines Skalierungsmodells für die Markt- und Meinungsforschung zurück.

Wir werden unser Modell auf folgendem Grundgedanken aufbauen. Die auf die Reihe der reellen Zahlen abgebildeten Mengen entsprechen der Menge aller möglichen Intensitäten der zu messenden Eigenschaft. Diese Intensitäten existieren nur an konkreten Objekten. Alle Relationen die zwischen den Intensitäten festgestellt werden, gehen auf Relationen zwischen den konkreten Objekten zurück. Diese konkreten Objekte können in unserem Fall einmal eine Reihe von Stimuli (zB. Warenproben, aber auch Spielsituationen, ferner Feststellungen und Behauptungen, kurz Items) zum andern Menschen sein. Die Aufgabe besteht nun darin, auf Grund von Beurteilungen

Ad 1

$$q(x, y) = q[\phi(x), \phi(y)] = q(a, b) = q(\tilde{a}, \tilde{b}) = q[\phi(\tilde{a}), \phi(\tilde{b})] = q(\lambda \cdot a, \lambda \cdot b)$$

das heißt aber

$$q(x, y) = q'(x - y)$$

Damit ist die obige Schlußfolgerung bewiesen. Für $x \prec y$ und $w(a, b)$ lassen sich ähnliche Beweise durchführen.

der in den Stimuli enthaltenen Intensitäten der zu messenden Eigenschaft zu skalieren. Die auf diese Weise aufgestellte Skala dient dann dazu, die in den Befragten enthaltenen Intensitäten der Eigenschaft zu messen. Da den Stimuli sowohl Rangziffern wie metrische Skalenwerte zugeordnet werden können, kann auch die Skalierung der Befragten sowohl topologisch wie metrisch erfolgen. Unser Modell gliedert mithin in vier Hauptabschnitte:

1. Die topologische Skalierung der Stimuli
2. Die topologische Skalierung der Befragten
3. Die metrische Skalierung der Stimuli
4. Die metrische Skalierung der Befragten.

Daraus ergibt sich der Aufbau der Arbeit: Nachdem wir im ersten Kapitel die Grundgedanken und die Problematik des Modells eingehender dargestellt haben, werden wir im zweiten Kapitel die Konstruktion einer topologischen Skala behandeln. Zuerst diskutieren wir dabei die Auswahl und Skalierung der Stimuli, dann die Skalierung der Befragten. Genauso gehen wir im dritten Kapitel vor, in dem die Konstruktion einer metrischen Skala für Meinungen und Einstellungen dargestellt wird. Im letzten Kapitel schließlich geben wir eine Übersicht über das gesamte Modell und zeigen Wege für seine mögliche Erweiterung auf den Fall mehrdimensionaler Eigenschaften.

1. Kapitel: Problemstellung und Überblick

=====

1. Definitionen

a. Skalierungstechnik ist die Entwicklung, Kombination und Anwendung von Methoden elementarstatistischen Messens mit dem Ziel, manifeste Bewusstseinsinhalte zu quantifizieren. Was bedeutet das im einzelnen ?

Wenn wir von Methoden elementarstatistischen Messens sprechen, so meinen wir damit nicht "Messen" im Sinne der Kollektivmaßlehre. Dort versteht man darunter die Berechnung einer oder mehrerer Maßzahlen, die einen komplexen Sachverhalt charakterisieren sollen.¹⁾ Also etwa eines Index. Am Beispiel des Index der Verbraucherpreise: Wir haben ein in

1) Anderson, C.: Probleme der statistischen Methodenlehre in den Sozialwissenschaften, Würzburg, 1957, S.19 ff.
Koller, S.: Zur Problematik des statistischen Messens, Allg.Stat.Arch. (40), 1956, S.316-340.
Pfanzagl, J.: a.a.O., 1959, S.9.
Winkler, W.: Statistische Maße, die ungleich messen, Allg.Stat.Archiv (40), 1956, S.297-315.

seiner Struktur und Qualität schwankendes Warenangebot, es liegen verschiedene Preiserhöhungen und Preissenkungen vor, zum Teil auf Grund dieser Preisveränderungen, zum Teil auf Grund anderer Faktoren verschiebt sich die Nachfrage - kurz und gut, der Tatbestand ist äußerst verwickelt. Dem Statistiker ist nun die Aufgabe gestellt, eine Maßzahl zu finden, die z.B. für im Hinblick auf Preissteigerungen erhobenen Lohnforderungen eine objektive Verhandlungsgrundlage bilden kann.

Mit dieser Art des Messens, die etwa Campbells "measurement of B magnitudes" oder "measurement by fiat" entspricht,¹⁾ beschäftigen wir uns nicht. Sie beruht auf einer entweder intuitiv oder auf Grund eines Modells oder Naturgesetzes geforderten Beziehung zwischen bestimmten Beobachtungen und dem Untersuchungsobjekt. Sie ist weder "elementar" noch "exakt". Das besagt nicht, daß diese Art des Messens in der Markt- und Meinungsforschung nicht angewendet werden könnte. In der Tat wird sie es gern und häufig. So kann man beispielsweise den Gottesdienstbesuch als Index für die "Einstellung zur Kirche" benutzen. Aber eine exakte Quantifizierung, wie sie von der Skalierungstechnik angestrebt wird, läßt sich dadurch natürlich nicht erzielen.

"Messen" wollen wir hier vielmehr in einem ursprünglicheren, direkteren Sinn. Nicht die Raffung bestehender Zahlen interessiert uns, sondern das Zustandekommen dieser Zahlen selbst. Sozusagen die Uraufschreibungen. Oder, wieder in Campbells Terminologie "measurement of A magnitudes" oder "fundamental measurement".²⁾ Dieses Messen setzt nicht eine

1) Campbell, N.R.: An account of the Principles of Measurement and Calculation, London, 1928.

2) Campbell, N.R.: a.a.O., 1928.

Messung anderer Variabler voraus, die zugeordneten Zahlen repräsentieren das Meßobjekt direkt. Genauer: die Menge M des Meßobjekts wird direkt auf die Menge der reellen Zahlen abgebildet, so, daß die Struktur der Bildmenge isomorph zur Struktur von M ist. Dieses direkte oder "fundamentale" Messen ist unser Ziel. So gesehen ist also Skalierungstechnik kein Pendant zur Konstruktion von Indices, sondern zur Konstruktion eines Metermaßes, einer Waage, eines Ampèremeter usw. Der so konstruierte Maßstab kann dabei durchaus, wie wir sehen werden, metrischen Charakter haben.¹⁾

b. Das Objekt unserer Messung sind "manifeste Bewußtseinsinhalte". Ein Ausdruck, der dem Statistiker zuerst nicht allzuviel sagt. In der empirischen Sozialforschung versteht man darunter im allgemeinen Meinungen und Einstellungen, das heißt "sedimentierte und bis zu einem gewissen Grad generalisierte Anschauungsweisen, Reaktionsformen, Haltungen und Ideologien." Soweit die Definition, wie sie von der empirischen Sozialforschung gegeben wird.²⁾

-
- 1) Die Frage, ob in der Psychologie überhaupt eine metrische Messung möglich sei, wurde lange diskutiert. Siehe z.B.: Bergmann, G. and K.W. Spence: The logic of psychological measurement, Psych.Rev. (51), 1944, S.1-24. Comrey, A.L.: An operational approach to some problems in psychological measurement, Psych.Rev. (57), 1950, S.217-228. McGregor, D.: Scientific measurement and psychology, Psych. Rev. (42), 1935, S.246-266. Stevens, S.S.: Mathematics, measurement and psychophysics, in Handbook of Experimental Psychology, New York, 1951, S.1-49. Die Skepsis entsprang vor allem der Tatsache, daß in der Psychologie additive Merkmale völlig fehlen, andererseits beim fundamentalen Messen physischer Eigenschaften stets die Forderung nach Erfüllung des Additivitätsaxioms erhoben wurde (die Messung nicht-additiver Eigenschaften, wie Temperatur, oder Dichte, erfolgt hier über abgeleitete Maße!). Neuere Arbeiten zeigen jedoch, daß zur Konstruktion einer metrischen Skala wesentlich schwächere Voraussetzungen ausreichen, als die der Additivität. Vgl. vor allem Pfanzagl, J.: a.a.O., 1959.
- 2) Artikel "Empirische Sozialforschung", Handwörterbuch der Sozialwissenschaften, Band 9, S. 419-435.

So deskriptiv diese Definition für den Soziologen sein mag, wir können auch damit vorerst nicht allzuviel anfangen. Versuchen wir unser Meßobjekt von einer anderen Seite her zu bestimmen.

Gegenstand jeder Beschreibung (und damit auch jeder exakten Beschreibung, ist gleich Messung) sind Eigenschaften. Wenn wir ein Grundstück beschreiben wollen, so geben wir seine Länge und Breite an. Ein Ziegelstein etwa wird durch Länge, Breite, Höhe, Gewicht, Farbe, Porosität beschrieben usw. Genauso beim Menschen, die ersten Seiten jedes Passes zeigen den Versuch, seinen Inhaber an Hand einiger seiner Eigenschaften wie Größe, Gewicht, Haar und Augenfarbe, Geschlecht und Alter zu beschreiben. Skalierung in einem weiteren Sinne ist beinahe schon das Henry'sche oder Bertillonsche Klassifikationssystem.¹⁾

Schon diese Beispiele legen eine generelle Einteilung der Eigenschaften nahe. Einmal in qualitative, zum andern in quantitative. Für die meisten Eigenschaften ist allerdings diese Einteilung willkürlich, oder besser, nur unabhängig vom jeweiligen Stand der Meßtechnik. Solange keine Waage existierte, konnte man das Gewicht nur als "schwer", "mittel" oder "leicht" angeben: es war also eine qualitative Eigenschaft. Jetzt ist es eine quantitative. Da es unser Ziel ist, solche "Waagen" zu entwickeln, also qualitative in quantitative Eigenschaften zu überführen, kurz, zu quantifizieren, können wir das Objekt unserer Untersuchung schon einmal näher begreifen: es sind qualitative Eigenschaften.

1) Sir Edward Richard Henry entwickelte 1901 das "Henry system of classification", das es erlaubt, auf Grund acht verschiedener Merkmale der Fingerabdrücke deren Inhaber nach einer bestimmten Formel zu klassifizieren. Das weniger bekannte Bertillon-System versucht ähnliches an Hand von Messungen der Knochenstruktur. Die Träger der Eigenschaften werden also auf einer Nominalskala "gemessen".

Die qualitativen Eigenschaften können jetzt wieder prinzipiell meßbar oder nicht meßbar sein. Meßbar sind sie dann, wenn sie kontinuierlich, das heißt, unbegrenzt unterteilbar sind. Diskontinuierliche Eigenschaften sind nicht meßbar. Körpergröße wäre ein Beispiel für eine kontinuierliche Eigenschaft "Geschlecht" eines für eine diskontinuierliche. Ein Metermaß kann man beliebig unterteilen, "männlich" und "weiblich" nicht. Es ist leicht einzusehen, daß daher eine Messung diskontinuierlicher Eigenschaften unmöglich und sinnlos ist. Sie lassen sich nur abzählen. Alle kontinuierlichen Eigenschaften sind theoretisch durchaus meßbar. Damit haben wir eine zweite Begrenzung unseres Untersuchungsobjekts: wir wollen qualitative, kontinuierliche Eigenschaften messen.

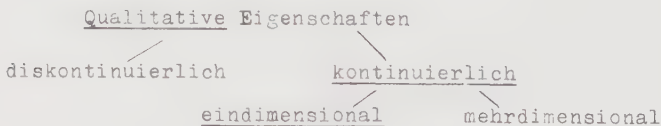
Noch ein drittes Einteilungskriterium ist denkbar: das Kontinuum dieser Eigenschaften kann ein- oder mehrdimensional sein. Eindimensional, das heißt, jede Anordnung von Objekten, die eine bestimmte Menge der betreffenden Eigenschaft besitzen, muß in Bezug auf diese Eigenschaft invariant sein. In Bezug auf ihre Körpergröße z.B. lassen sich, sagen wir, zehn Menschen nur auf eine ganz bestimmte Art und Weise anordnen. Anders wäre es, wollten wir zehn rote Farbkarten nach ihrer Tönung ordnen. Je nach dem ob man mehr die Leuchtkraft, die Sättigkeit usw. der Farbe berücksichtigt, wird die Anordnung verschieden sein. Die Eigenschaft "Farbe" besitzt also ein mehrdimensionales Kontinuum. Durch ausreichende Spezifikation läßt es sich allerdings auf mehrere eindimensionale zurückführen. Wir werden uns in dieser Arbeit hauptsächlich mit

eindimensionalen Eigenschaften beschäftigen. Auf die „öglichkeiten mehrdimensional zu messen werden wir indessen hinweisen. Vorerst können wir jedoch sagen: Untersuchungsobjekt sind qualitative, kontinuierliche, eindimensionale Eigenschaften.

Noch eine letzte Unterteilung ist nützlich, da wir es in der empirischen Sozialforschung mit menschlichen Eigenschaften zu tun haben, ist es vielleicht noch nützlich zwischen physischen und psychischen Eigenschaften zu trennen. Offensichtlich gehören "manifeste Bewußtseinsinhalte" zu den psychischen Eigenschaften.

Fassen wir zusammen: Das Objekt unserer Messung sind Meinungen und Einstellungen. Sie gehören zu den qualitativen, psychischen Eigenschaften und sind dann mit den hier vorgeführten Methoden meßbar, wenn sie kontinuierlich und eindimensional sind.

Schaubild 1



Damit haben wir sowohl "elementarstatistisches Messen" als auch das Objekt unserer Messung definiert. Die etwas ausführliche Behandlung der "Eigenschaften" war notwendig. Der Begriff des Kontinuum und der Ein- und Mehrdimensionalität ist wesentlich und wird uns durch die ganze Arbeit begleiten.

2. Methoden der Messung

a. Angenommen, wir wollten die Verteilung einer ganz bestimmten Einstellung auf eine bestimmte Gruppe von Menschen messen. Etwa die "Einstellung der deutschen Wähler zur Integration Westeuropas". Weiterhin angenommen, diese Einstellung sei in der von uns geforderten Form vorhanden, das heißt, sie besitze ein eindimensionales von negativ über neutral bis positiv laufendes Kontinuum.

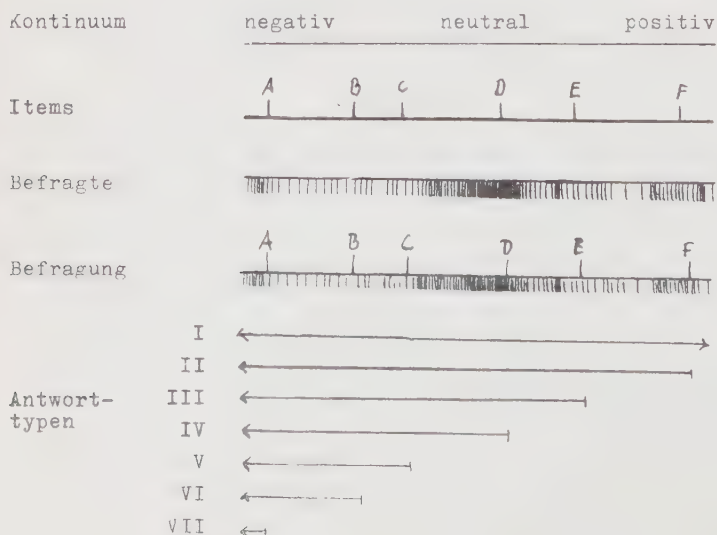
Die einfachste und nächstliegende Möglichkeit wäre nun, eine repräsentative Teilgesamtheit von Wählern zu fragen: "Sind Sie für oder gegen eine Integration Westeuropas?" Indessen, diese Möglichkeit interessiert uns hier nicht. Ganz abgesehen davon, daß das Wort "Integration" für eine solche Befragung viel zu schwierig und mißverständlich wäre, wird eine derartige, direkte Frage wahrscheinlich verzerrte Ergebnisse liefern. Vor allem aber reicht sie natürlich nicht für eine echte "Messung" der betreffenden Einstellung, bzw. ihrer Verteilung auf die Befragten aus.¹⁾

Was läge also näher, als durch eine ganze Reihe von Fragen oder auch Behauptungen, Feststellungen (wir wollen sie im folgenden kurz Items nennen) das Kontinuum zu unterteilen. Etwa in der Art: "Sind Sie für oder gegen gute Beziehungen zwischen den westeuropäischen Staaten?" "Sind Sie für oder gegen die Bildung eines westeuropäischen Bundesstaates mit

1) Edwards, A.L.: Techniques of Attitude Scale Construction, New York, 1957, S.3 f.

Deutschland als Mitglied?" "Sind Sie für oder gegen die Bildung eines westeuropäischen Gesamtstaates, in dem Deutschland voll und ganz aufgeht?" usw. Alle diese Items repräsentieren einen bestimmten Grad an positiver oder negativer Einstellung zur Integration Westeuropas und lassen sich auf dem Kontinuum dieser Einstellung entweder rangmäßig oder auch - wie wir sehen werden - metrisch anordnen. Genauso wie diese Items hat aber auch jeder Befragte, entsprechend seiner Einstellung, eine bestimmte Position auf dem Kontinuum. Das heißt aber: legt man die einzelnen Items den einzelnen Befragten vor, so wird man - unter bestimmten Voraussetzungen - aus der Verteilung der Antworten auf die Verteilung der "Einstellung zur Integration Westeuropas" auf die Gesamtheit der deutschen Wähler schließen können. Schematisch dargestellt sieht das etwa so aus:

Schaubild 2



b. Wenn nun die vorhin aufgestellte Hypothese, daß die "Einstellung zur Integration Westeuropas" ein eindimensionales Kontinuum besitzt, richtig ist, so werden sich für die sechs Items sieben typische Befragtengruppen herausbilden: zuerst diejenigen, die extrem positiv zu dem Untersuchungsobjekt stehen und alle Items bejaht haben, dann diejenigen, die - nicht ganz so extrem in ihrer Einstellung - alle Items bis auf das letzte positiv beantwortet haben usw., herunter bis zu den stark nationalistisch denkenden Befragten, die alle Items verneint haben.

Schaubild 3

	Fragen						Σ
	A	B	C	D	E	F	
I	+	+	+	+	+	+	n_1
II	+	+	+	+	+	-	n_2
III	+	+	+	+	-	-	n_3
IV	+	+	+	-	-	-	n_4
V	+	+	-	-	-	-	n_5
VI	+	-	-	-	-	-	n_6
VII	-	-	-	-	-	-	n_7

Damit aber haben wir zweierlei erreicht. Erstens gibt uns das Ausmaß, mit dem sich die empirischen Daten diesem Modell anpassen, einen Test auf die Richtigkeit unserer Hypothesen. Wenn in signifikantem Ausmaß Antworttypen wie (++++-+), (+++--+), oder gar (---++) auftauchen, ist die Hypothese eines eindimensionalen Kontinuum offensichtlich falsch. Zweitens haben wir in der Verteilung der Befragten auf die einzelnen Antworttypen Näherungswerte für deren Verteilung auf dem Kontinuum.¹⁾

1) Siehe dazu: L. Guttman: The quantification of a class of attributes: A theory and method of scale construction. In Horst, Paul et al: The Prediction of Personal Adjustment, New York, 1941.

c. Allerdings unter der Voraussetzung, daß uns der Skalenswert der einzelnen Fragen oder Items bekannt ist. Das heißt, daß die einzelnen Fragen, die wir quasi als Maßstab benutzt haben, auf einer metrischen Skala gegeben und nicht nur topologisch, also rangmäßig geordnet sind. Die Antwortnauigkeiten auf einer Rangskala lassen keine eindeutigen Rückschlüsse auf die Verteilung der Befragten in Bezug auf die untersuchte Eigenschaft zu. Sie geben bestenfalls einen groben Schätzwert. 35 v.H. der Befragten in Antwortgruppe III würde hier nur besagen, daß 35 v.H. der befragten Wähler eine Einstellung hat, die ungefähr zwischen den in C und D formulierten Grenzen liegt - wie positiv oder negativ sie jedoch ist, kann daraus nicht entnommen werden.

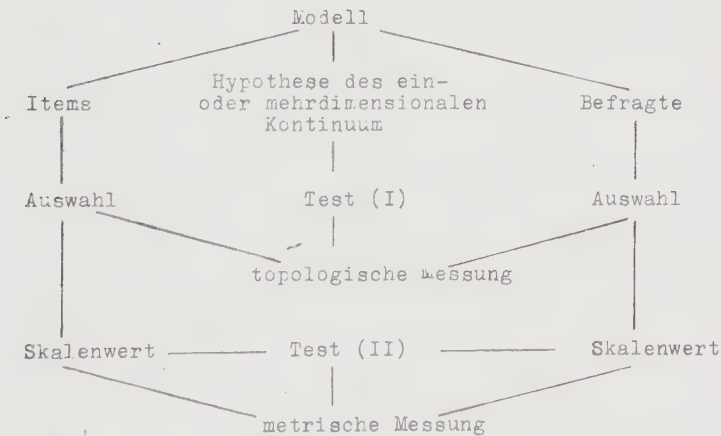
Es entsteht also die Frage: Gibt es Methoden, die die Konstruktion einer metrischen Skala für die Items erlauben? A priori erscheint es fragwürdig, einzelnen Items bestimmte metrische Werte zuzuordnen zu wollen. Etwa: B ist von C 2,7 Maßeinheiten, C von D 4,8 Maßeinheiten entfernt. Mehr als eine rangmäßige Anordnung läßt sich - so sieht es aus - nicht erzielen. Wir werden jedoch sehen, daß auf dem Gebiet der Experimentalpsychologie und Soziometrie Methoden entwickelt wurden, die eine metrische Messung erlauben und deren analoge Anwendung hier möglich ist.

d. Fassen wir zusammen: Das Modell, dessen wir uns bei der Quantifizierung von Einstellungen und Meinungen bedienen wollen, besteht praktisch aus drei Stufen - der Konstruktion

eines Maßstabs mit Hilfe des Items, das Anlegen dieses Maßstabs an die zu untersuchende Gesamtheit, das Testen der zugrunde liegenden Hypothesen.

Dabei ergibt sich erstens das Problem der richtigen Auswahl, sowohl der Items wie der Befragten; zweitens das des Tests und - wenn eine metrische Messung beabsichtigt ist - ; drittens das der Bestimmung des Skalenwerts, also der genauen Position der Items und Befragten auf dem Kontinuum. Die Bausteine des Modells sind in Schaubild 4 schematisch dargestellt:

Schaubild 4



Die Bezeichnung Test(I) und Test (II) bezieht sich auf zwei Möglichkeiten des Testens. Die erste ist die bereits in ihren Grundzügen erwähnte Methode, die Häufigkeit des Auftretens atypischer Antwortkombinationen als Test zu verwenden.

den. Dieser Test kann natürlich nur gewisse Daumenregeln für Akzeptierung oder Zurückweisung der Hypothese liefern. Für eine metrische Messung indessen genügt er nicht. Wir werden daher im zweiten Fall andere, exaktere Methoden anwenden.

3. Überblick über den Aufbau der Arbeit

Aus dieser Skizze der Grundgedanken unseres Modells ergeben sich schon die Aufgaben und Probleme, die uns erwarten, und damit die Gliederung der kommenden Abschnitte. Im nächsten Kapitel werden wir die Methoden schildern, die eine topologische Messung ermöglichen. Dabei ergibt sich als erstes das Problem der Auswahl, sowohl der Items als auch der Befragten. Eine große Zahl von Items, die sich auf das Kontinuum beziehen und es in irgendeiner Weise abteilen, ist möglich. Welche soll man nun benutzen, welche Kriterien sind denkbar, an Hand derer wir die relevanten unter ihnen herausfinden können? Auch die Auswahl der Befragten ist ein Problem - da es jedoch in der Stichprobentheorie ausgiebig behandelt wird, brauchen wir es hier nur kurz zu streifen. Dann kommen wir zur "Meinungsumfrage" selbst. Also zur (vorerst topologischen) Bestimmung der Verteilung der untersuchten Eigenschaft auf die Befragten. Zugleich stellt sich uns die Frage des Tests. Sind die zugrunde liegenden Hypothesen richtig? Kann man die Einstellung als "skalierbar" betrachten? Die vorhin erwähnte Testmethode erscheint relativ unproblematisch - müssen, bei ihrer genaueren Betrachtung und Anwendung

werden sich einige Schwierigkeiten ergeben. Danach müssen wir Station machen, die Ergebnisse der topologischen Messung resümieren und ihren Aussagewert prüfen.

Im darauffolgenden Kapitel gehen wir einen Schritt weiter und diskutieren Möglichkeit, Methode und Ergebnis einer metrischen Messung. Wieder geht es zuerst um die Konstruktion des Maßstabes - hier die Bestimmung des Skalenwerts der Items, oder wie wir dann allgemeiner sagen werden, der Stimuli. Dann befassen wir uns mit der Messung der Verteilung, also der (jetzt metrischen) Skalierung der Befragten. Der Test auf die notwendige Annahme der Eindimensionalität kann diesmal schon auf der ersten Stufe durchgeführt werden.

Im letzten Kapitel geben wir zuerst einen Überblick über unser nun vollständiges Gesamtmodell und wenden uns dann der Frage zu: Was geschieht, wenn wir die Abweichungen der empirischen Daten vom Modell nicht mehr als zufällig behandeln können, wenn wir also die allem zugrunde liegende Hypothese der Eindimensionalität zurückweisen müssen? Wir werden sehen, daß dann noch die theoretische Möglichkeit mehrdimensionaler Modelle existiert, wenn auch eine praktische Anwendung vorerst noch kaum möglich ist.

2. Kapitel: Die Konstruktion einer topologischen Skala für Meinungen und Einstellungen

Unter einer topologischen Skala wollen wir im folgenden, in Anlehnung an Carnap¹⁾ und Pfanzagl²⁾ eine Skala verstehen, die nur eine Ordnung wiedergibt (Rangskala, "ordinal scale"). Für ihre Konstruktion ist, gemäß dem oben skizzierten Modell, zunächst die richtige Auswahl der Items und Befragten notwendig. Sie wird in Abschnitt A behandelt werden. Es folgt in Abschnitt B die Darstellung der Methode der eigentlichen Messung und des analytischen Vorgehens beim Testen auf Ein-dimensionalität. Abschnitt C bringt schließlich eine Zusammenfassung und Diskussion der erzielten Ergebnisse.

A. Die Auswahl der Items.

1. Begriff und Aufbau des Item

a. Bei der Skalierung von Meinungen und Einstellungen haben wir ein doppeltes Auswahlproblem zu lösen: einmal die Auswahl der "geeignetsten" Items, zum andern die Auswahl einer repräsentativen Teilgesamtheit von Befragten. Hier nun zuerst zur Auswahl der Items.

Items, so hatten wir weiter oben gesagt, können Fragen, Aussagen, Teststellungen usw. sein. Ganz allgemein werden wir sagen, was geeignet ist, die

Befragten in Bezug auf eine bestimmte Meinung oder Einstellung in zwei oder mehr sich gegenseitig anschließende Kategorien zu unterteilen.¹⁾ Alles also, was das Kontinuum in geeigneter Weise abteilt.

Für die Form der Items sind eine Reihe von Möglichkeiten denkbar. Sie können kategorisch oder dichotym, monoton oder punktuell sein. An einem Beispiel: wollten wir - völlig absurderweise - auf Grund einer Befragung die Verteilung der Eigenschaft "Körpergröße" auf eine Gruppe von Menschen ermitteln, so könnten wir das dort unzweifelhaft vorhandene eindimensionale (physische) Kontinuum durch folgende Items abteilen.

1. Beispiel: Die richtige Kategorie ist anzukreuzen:

Meine Größe ist

- a) über 1,80 m
- b) zwischen 1,70 und 1,80 m
- c) zwischen 1,60 und 1,70 m
- d) unter 1,60 m

Dies wäre ein Beispiel für ein kategorisches Item. Es teilt das Kontinuum der Eigenschaft "Körpergröße" in vier verschiedene Kategorien, zwei offene (a,d) und zwei geschlossene (b,c).

2. Beispiel: Sind Sie

- 1. Unter 2.20 m?
- 2. Unter 2.00 m?
- 3. Unter 1.80 m?
- 4. Unter 1.70 m?
- 5. Unter 1.60 m groß?

ja	nein

Hier haben wir fünf dichotyme Items. Für jedes einzelne sind nur zwei Antwortkategorien möglich. Im ersten Beispiel hatten

1) Torgerson, W.S.: Theory and Methods of Scaling, New York, 1958.

wir ein Item mit vier Antwortkategorien. Ferner sind die Items in diesem Beispiel monoton. Das heißt, alle Befragten, die sich unterhalb der vom Item gezogenen Trennungslinie befinden, werden mit "ja", alle die sich oberhalb davon befinden mit "nein" antworten. Items dieser Art werden auch mit "increasing probability items"¹⁾ oder "cumulative items"²⁾ bezeichnet. Die hier verwendete Bezeichnung lehnt sich an den Ausdruck "monotone items" von Coombs an.³⁾

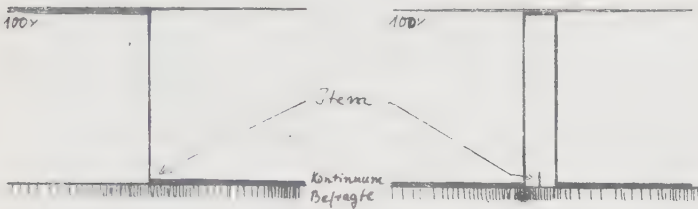
<u>3. Beispiel:</u> Meine Größe ist	ja	nein
1. Um 2.00 m		
2. Um 1.90 m		
3. Um 1.80 m		
4. Um 1.70 m		
5. Um 1.60 m		
6. Um 1.50 m		

Auch diese sechs Items sind ohne Zweifel dichotym, sie besitzen nur zwei Antwortkategorien. Indessen, jeder Befragte wird nur eines der Items mit "ja" beantworten. Sowohl diejenigen, die sich oberhalb seiner Position auf dem Kontinuum, als auch diejenigen, die sich unterhalb davon befinden, werden mit "nein" antworten. Die Wahrscheinlichkeit einer positiven Antwort wird so verlaufen, wie Schaubild 5 (b) zeigt, wogegen bei monotonen Items die Wahrscheinlichkeitslinie so, wie in

-
- 1) Thurstone, L.L. and E.J. Chave: The Measurement of Attitude, Chicago, 1929.
 - 2) Loevinger, J.: The technic of homogeneous tests compared with some aspects of "scale analysis" and factor analysis, Psych.Bull. (45), 1948, S.507-5

5 (a) gegeben ist.

Schaubild 5



Wahrscheinlichkeit
einer positiven Antwort
auf ein monotones Item

Wahrscheinlichkeit
einer positiven Antwort
auf ein Punktitem

Items, wiesie uns das dritte Beispiel zeigt, nennen wir mit Mosteller "Punktitems".¹⁾ Coombs nennt sie "nonmonotone", Thurstone/Chave "maximum probability" Items.²⁾

Das Beispiel läßt sich leicht auf den Fall der Skalierung von Einstellungen und Meinungen übertragen. Auch hier sind diese drei Itemtypen möglich: kategorische und dichotyme, letztere wieder monoton oder punktuell. Jeder Itemtypus erfordert ein leicht verändertes Modell. Wir wollen indessen im folgenden, der Einfachheit halber, unter "Items" immer dichotym, monotone verstehen. Sollte in manchen Fällen eine andere Formulierung der Items notwendig sein, so werden wir darauf hinweisen.

2. Kriterien für die Auswahl der Items

Soviel zum formalen Aufbau der Items. Nun zu unserem eigentlichen Problem: der Auswahl. Was für Anforderungen müssen

1) Mosteller, F.: Remarks on the theory of paired comparison, I, II, III. Psychometrika (16), 1951, S. 5-11, S. 223-226, S. 207-218.

2) Coombs und Thurstone and Chave, a.a.O., 1958 bzw. 1929.

wir an die Items stellen? Welche Auswahlkriterien können wir anwenden?

Ohne Zweifel ist in Bezug auf eine bestimmte Eigenschaft oder Meinung eine Vielzahl von Behauptungen, Fragen, kurz Items denkbar. Wir wissen jedoch nicht, ob diese sich wirklich eindeutig auf die betreffende Eigenschaft beziehen, wo ungefähr sie sich auf dem Kontinuum befinden usw. Kurzum, was wir brauchen sind Kriterien und Methoden, die es uns erlauben, aus der Schar der möglichen Items die für uns geeignetste herauszufinden.

Sicher, es existiert eine Reihe von Regeln für die richtige Formulierung der Items: die Sprache soll einfach, klar und direkt sein, keine doppelte Verneinung, keine schwierigen Fremdworte, keine langen Nebensätze. Die Behauptungen und Feststellungen sollen eindeutig sein, sich auf die Gegenwart beziehen usw.¹⁾. Mit dieser Seite des Auswahlproblems wollen wir uns hier nicht beschäftigen. Wir nehmen an, jedes der uns vorliegenden Items sei gut formuliert. Was ist nun das Ziel der Auswahl? Was heißt "geeignete Items"? Erinnern wir uns an unser Modell. Aus den Items wollen wir den Maßstab, die Skala konstruieren, den wir dann an die Befragten anlegen, um die Verteilung der Eigenschaft

1) Kriterien dieser Art wurden hauptsächlich von K.A. Wang: Suggested criteria for writing attitude statements, Journ. soc. Psychol., 1932 und S.L. Payne: The art of asking questions, Princeton, N.J., 1951, vorgeschlagen. Eine Zusammenfassung findet sich bei Edwards, a.a.O., 1957, S.13 f. Siehe auch R. König (Hrsg.), Praktische Sozialforschung I. Das Interview, Formen - Technik - Auswertung, Köln, 1957.

"Einstellung zu ..." zu messen. Das heißt aber sie müssen folgenden Anforderungen genügen:

- a) sie müssen den relevanten Teil des Kontinuum abdecken
- b) sie müssen auf diesem möglichst gleichmäßig verteilt sein
- c) sie müssen sich eindeutig auf das Kontinuum beziehen und gut diskriminieren, das heißt also, das Kontinuum möglichen scharf abteilen.

Was wir demnach brauchen, ist eine Technik, die es uns erlaubt, an Hand statistischer Kriterien die vorgeschlagenen Items nach diesen Gesichtspunkten quasi zu "sieben". Zwei Methoden, die hin und wieder vorgeschlagen werden, scheiden damit aus: Einmal die Zufallsstichprobe, wir wollen ja keine repräsentative Teilgesamtheit der fehlerhaften Items, sondern eine Auslese der geeignetsten. Zum andern der Vorschlag Guttman's, diese Auslese auf Grund von Intuition und Erfahrung zu treffen,¹⁾ unsere Kriterien sollen ja "statistisch" sein. Was also bleibt uns?

3. Auswahlmethoden

a. Wir wollen im folgenden eine Auswahltechnik entwickeln, die uns in drei statistischen Maßen, α , α_1 und α_2 , eine relativ exakte Auswahl der den obigen Kriterien am besten entsprechenden aus der Gesamtzahl der vorgeschlagenen Items ermöglicht. Der mit der Materie vertraute Leser wird in der hier vorgeschlagenen Technik Gedanken Thurstones, Likerts, Edwards und Guttman's finden.²⁾ Die von diesen Autoren vorgeschlagenen Metho-

1) Guttman, L., a.a.O., 1941.

2) Thurstone, L.L. and E.J. Chave, a.a.O., 1929.

Likert, R.: A technique for the measurement of attitudes, Arch. Psych., Nr. 140, 1932.

Edwards, A.L. and F.P. Kilpatrick: A technique for the construction of attitude scales. J. exp. soc. psych. (70), 1948.

den der Itemanordnung oder Itemanalyse haben jede für sich durchaus ihre Vorzüge. Thurstones Modell der "equal appearing intervals", das hier nur rangmäßigen Anordnung und groben Klassifizierung der Itemgesamtheit benutzt wird, kann mit einiger Phantasie und einigem Optimismus zu einem primitiven Skalierungsmodell überhaupt ausgebaut werden (es ist nicht sicher, ob Thurstone selbst das uneingeschränkt getan hätte).¹⁾ Indessen, aus verschiedenen Gründen²⁾ wollen wir hier diese Technik wesentlich bescheidener handhaben, nur als Hilfsmittel bei der Auswahl, sie aber zu diesem Zweck noch durch die Itemanalyse Likerts und Edwards ergänzen. Wohlgedenkt also, wir befinden uns erst im Abschnitt "Auswahl der Items": die im folgenden dargestellte Technik ist also kein Skalierungsmodell, sie liefert nur Kriterien für die Auswahl der später zu skalierenden Items.

b. Der Grundgedanke dieser Auswahltechnik. Praktisch gehen wir in zwei Stufen vor. Zuerst werden alle vorgeschlagenen Items einer Beurteilergruppe vorgelegt, deren Mitglieder in bestimmter Form eine direkte, quantitative Schätzung der Position der Stimuli auf dem Kontinuum machen sollen. Ursprung und Einheit der Skala sind dabei definiert.

Man nimmt an, daß sich mehrere Schätzungen eines Beurteilers oder je eine Schätzung mehrerer Beurteiler um die wahre Position des Stimulus auf dem Kontinuum verteilen werden.

1) Torgerson, W.S., a.a.O., 1958, S.61 ff.

2) Ein Hauptargument gegen den Gebrauch von Thurstones Methode der "equal appearing intervals" als Skalierungsmodell ist, daß durch die Anordnung des Experiments den Daten praktisch Transivität aufgezwungen wird, das heißt aber, kein Test darauf, ob die Skala wirklich sinnvoll ist, möglich. In der Tat können völlig unzusammenhängende Items auf diese Art formal skaliert werden.

Das heißt nun aber, ein Mittelwert für die verschiedenen Plazierungen der Items kann eine ungefähre Vorstellung von deren wahrer Position auf dem Kontinuum liefern. Ein solcher Mittelwert wird unser Kriterium $\mathcal{A}$, sein. An Hand dieses Kriteriums können wir schon einmal eine versuchsweise rangmäßige Anordnung der vorgeschlagenen Items auf dem Kontinuum vornehmen. Darüber hinaus sehen wir auch ungefähr, wo sich die Items besonders "ballen" bzw. wo "dünne Stellen" bestehen.

Weiterhin wird ein Maß für die Streuung der einzelnen Beurteilungen eine Vorstellung von dem Grad der Übereinstimmung geben, die bei der Beurteilung geherrscht hat. Dieses Streuungsmaß ist unser Kriterium $\mathcal{L}_1$. Auf Grund von $\mathcal{A}$, können wir schon einen großen Teil der vorgeschlagenen Items ausscheiden. Dabei müssen wir allerdings immer auf $\mathcal{L}$, achten, d.h. eine möglichst gleichmäßige Verteilung der restlichen Items auf dem Kontinuum anstreben.

Die Items werden nun einer zweiten Beurteilung unterzogen. Die Aufgabe der Beurteiler ist jetzt indessen eine andere. Sie sollen keine Schätzung der Position des Items auf dem Kontinuum mehr abgeben, sondern den Grad ihrer persönlichen Einstellung zu der in dem Item geäußerten Behauptung oder Feststellung kenntlich machen - etwa durch Ankreuzen einer von sechs oder vier Antwortklassen (z.B. "stark dagegen, dagegen, etwas dagegen, etwas dafür, dafür, stark dafür"). Hier haben wir also sozusagen schon eine Art "Probe-Meinungsumfrage". Im ersten Fall sollte die Beurteilung nicht durch die persönliche Einstellung beeinflusst werden, im zweiten Fall ist gerade

diese es, die uns interessiert.¹⁾

Offensichtlich sind diejenigen Items, die eine starke Reaktion hervorzurufen vermögen, "stark dagegen" oder "stark dafür", besser geeignet das Kontinuum zu unterteilen. Unser Kriterium α_i wird ein Maß für diese Diskriminanzfähigkeit der Items geben. Damit werden wir aber in der Lage sein einen weiteren, großen Teil der Items auszuschneiden und zu der geplanten Zahl der "besten Items" zu gelangen.

Die Grundlage unserer Auswahltechnik baut, speziell in ihrer ersten Stufe, offensichtlich auf der Annahme auf, daß eine ausreichende Wiedernholung subjektiver Schätzungen zu einer objektiven Aussage über das betreffende Objekt führen kann. Daß diese Annahme berechtigt ist, ergibt sich aus zahlreichen Experimenten der Psychophysik.²⁾ Hier wurde versucht, die auf Grund subjektiver Schätzungen physischer Eigenschaften erhaltenen Skalen zu den objektiv gegebenen in Beziehung zu setzen. Also etwa die Schätzung mehrerer Steine nach ihrem Gewicht zu ihrem effektiven Gewicht. Die Ergebnisse zeigten eine lineare Beziehung zwischen der subjektiv, psychischen und der objektiv, physischen Skala. Allerdings wurden diese Experimente wesentlich raffinierter aufgebaut als unser einfaches Modell.³⁾ Aber, wie gesagt, dieses soll ja nicht der Skalie-

1) Torgerson (a.a.O., 1958) teilt nach diesem Kriterium sogar die einzelnen Skalierungsmethoden ein. Er unterscheidet den "judgment approach", bei dem es nur um eine Skalierung der Items geht, also nur Beurteilung, und den "response approach", bei dem Items und Befragte skaliert werden, also nicht Beurteilung sondern Bekanntgabe der Einstellung.

2) Thurstone, L.L.: Psychophysical Methods. In T.G. Andrews (ed.): Methods of Psychology, New York, 1948.

3) Wir werden später, bei Thurstones Methode der komparativen Schätzung solch ein "raffiniertes" Modell kennenlernen.

rung sondern nur der Auswahl dienen. Dafür ist es in der vorliegenden Form ausreichend.

Zur Auswahl und Größe der Beurteilergruppe läßt sich folgendes sagen: Sie sollte möglichst eine kleine Stichprobe aus der später zu untersuchenden Grundgesamtheit darstellen - etwa in der Art eines "pilot sample". Über den Umfang dieser Vorstichprobe lassen sich keine Regeln liefern. Thurstone und Chave¹⁾ benutzten bei der ersten Anwendung ihrer Methode der "equal appearing intervals" 300 Beurteiler für 130 Items. Andere sind mit weit weniger umfangreichen Stichproben ausgekommen. Edwards und Kenney²⁾ z.B. haben 129 Items von 72 Beurteilern evaluieren lassen, Uhrbrock³⁾ ließ 279 Items von zwei Gruppen à 50 Beurteilern skalieren und kam zu einer Korrelation von 0,99 zwischen beiden Ergebnissen. Eine Gruppe von 50 Beurteilern dürfte also durchaus genügen.

c. Der Aufbau des Experiments und der Auswertung des gewonnenen Datenmaterials. Bei der Berechnung von $\mathcal{L}_1$ und $\mathcal{L}_2$ wollen wir ungefähr auf dem gleichen Weg vorgehen, der von Thurstone und Chave bei der ersten Anwendung der Methode der "equal appearing intervals" eingeschlagen wurde. Jedes der vorgeschlagenen Items - es können 50, 100, 200 sein - wird auf eine Karte geschrieben. Diese Karten werden nun dem einzelnen Beurteiler vorgelegt, dessen Aufgabe es ist, sie so in eine Reihe von Häufchen zu sortieren, daß die Intervalle zwischen

1) Thurstone, L.L. and E.J. Chave: a.a.B., 1929.

2) Edwards, A.L. and K.C. Kenney: A comparison of the Thurstone and Likert techniques of attitude scale construction, Journ. appl. Psych. (30), 1946, S.72-83.

3) Uhrbrock, R.S.: Attitudes of 4430 employees, Journ.soc. Psych. (5), 1934, S.365-377.

den Häufchen subjektiv gleich sind. Dabei werden die Beurteiler speziell davor gewarnt, eine bestimmte Verteilung der Items anzunehmen.

Wichtig ist es, Ursprung und Einheit zu definieren. Wird es den Beurteilern selbst überlassen, ihre Bezugsgrößen zu wählen, so wird diese von Person zu Person variieren. Das heißt aber, in der im folgenden beschriebenen Berechnung von $\mathcal{L}_1$ und $\mathcal{L}_2$ werden die Beurteilungen, die auf Grund einer kleineren Maßeinheit getroffen wurden, stärker gewichtet werden als diejenigen, die auf Grund einer größeren gemacht wurden. Die Bestimmung von Ursprung und Einheit kann allerdings nicht dadurch geschehen, wie manchmal vorgeschlagen wird, daß man drei Punkte der Skala festlegt. Also etwa "Die Items sollen auf einer siebengeteilten Intervallskala, die von -3 über 0 zu +3 geht in Bezug auf den Grad an positiver oder negativer Einstellung zu ... geordnet werden. Dabei steht -3 für das negativste, +3 für das positivste und 0 für ein neutrales Item." Obwohl diese Aufgabe recht vernünftig scheint, kann sie undurchführbar sein. Die Bedingung, daß die Intervalle zwischen den einzelnen Häufchen gleich sind, kann nur dann erfüllt werden, wenn der Abstand zwischen dem negativsten Item und 0 genau gleich ist dem Abstand zwischen dem positivsten Item und 0. Das heißt, allgemein gesprochen, wenn wir Ursprung und Einheit festlegen wollen, so können wir nur zwei Punkte der Skala bestimmen. Also etwa die beiden Extreme.¹⁾

Die einzelnen Items werden nun den Beurteilern in zufälliger Folge vorgelegt. Eine bestimmte vorherige Anordnung

1) Siehe dazu auch Torgerson, W.S., a.a.O., 1958, S.65.

würde wahrscheinlich keinen Einfluß auf das Ergebnis haben, andererseits sollte man sich selbst dieser unwahrscheinlichen Gefahr nicht aussetzen, wenn sie so leicht zu umgehen ist.

Das so gewonnene Ausgangsmaterial besteht aus einer Reihe von "Noten", die von den einzelnen Beurteilern jedem Item gegeben wurden. Die Zahl der "Noten" wird im allgemeinen für jedes Item gleich sein. Die Daten lassen sich in Form einer Häufigkeitsmatrix darstellen:

	Intervalle			
	1	2	g	m+1
1	f_{11}	f_{12}	f_{1g}	$f_{1,m+1}$
2	f_{21}	f_{22}	f_{2g}	$f_{2,m+1}$
...				
j	f_{j1}	f_{j2}	f_{jg}	$f_{j,m+1}$
...				
n	f_{n1}	f_{n2}	f_{ng}	$f_{n,m+1}$

In den Zeilen ($j = 1, 2, \dots, n$) stehen die Items, in den Spalten ($g = 1, 2, \dots, m+1$) die Intervalle. Das Element f_{jg} gibt an, wie oft Item j in Intervall g eingeordnet wurde. Am einfachsten ist es, wenn man diese F-Matrix der einfachen Häufigkeiten in eine P-Matrix der Anteilshäufigkeiten umwandelt, so daß p_{jg} den Anteil der Male angibt, in denen Item j in Intervall g sortiert wurde.

Die Berechnung von $\bar{x}_j$ ist nun elementar. Es ist nichts anderes als ein Mittelwert, der den einzelnen Items zugeordneten

Noten. Wir wollen den Median verwenden. Er läßt sich auf die bekannte Weise entweder durch graphische oder numerische Interpolation gewinnen. Im Falle einer numerischen Interpolation können wir für $\mathcal{L}_j$ die Formel

$$(1) \quad \mathcal{L}_j = g + \left(\frac{0,5 - \sum_{i=1}^g p_{ji}}{p_{j,g+1}} \right)$$

angeben, wobei wir als Intervalleinheit 1 annehmen und mit g die untere Grenze des Intervalls bezeichnet, in das $\mathcal{L}_j$ fällt. Nun ist $\mathcal{L}_j$ noch kein wesentliches Kriterium für die Auswahl der Items. Es dient uns vorerst nur dazu, einen ersten Überblick über die Anordnung der Items auf dem Kontinuum zu bekommen, sowie sie wenigstens versuchsweise rangmäßig zu ordnen. Wichtiger ist $\mathcal{L}_j$ das Maß für die Streuung der Beurteilungen.

Auch hier wollen wir wieder das einfachste und unproblematischste Streuungsmaß benutzen: den Quartilabstand. Auch er wird auf dem hinreichend bekannten Wege entweder der graphischen oder numerischen Interpolation gewonnen. Im zweiten Fall könnten wir wieder schreiben:

$$(2) \quad \mathcal{L}_{2j} = Q_{75,j} - Q_{25,j}$$

$$(3) \quad Q_{75,j} = h + \left(\frac{0,75 - \sum_{i=1}^h p_{ji}}{p_{j,h+1}} \right)$$

$$(4) \quad Q_{25,j} = f + \left(\frac{0,25 - \sum_{i=1}^f p_{ji}}{p_{j,f+1}} \right)$$

wobei $Q_{75,j}$ und $Q_{25,j}$ die beiden Quartilen der Anteilshäufigkeiten von Item j darstellen, h und f respektive die unteren Grenzen der Intervalle, in die diese fallen.

Offensichtlich wird $\mathcal{L}_i$ relativ klein sein, wenn bei den Beurteilern große Übereinstimmung bei der "Benotung" der betreffenden Items geherrscht hat. Umgekehrt zeigt ein großes $\mathcal{L}_i$, daß man sich relativ uneinig war. Das heißt aber in diesem Fall, daß das betreffende Item in irgendeiner Weise ungeeignet ist, bei der späteren Einstellungsskalierung das Kontinuum scharf abzutellen. Wahrscheinlich wird man annehmen können, daß es doppeldeutig ist, also, daß es von den Beurteilern in mehr als einer Weise interpretiert wurde. Indessen, was auch immer der Grund für die hohen Werte von $\mathcal{L}_i$ sein mag, die betreffenden Items sind für unser weiteres Vorgehen ungeeignet und können ausgeschieden werden. Bei welchem Wert von $\mathcal{L}_i$ man dabei die Grenze zieht, hängt ganz davon ab, wieviel Items man als Ausgangsmaterial hatte und wieviele man am Schluß für die eigentliche Skalierung behalten will. Im allgemeinen wird man 50% der Items ausscheiden. Das heißt also, der Median aller $\mathcal{L}_i$ Werte wird als Kriterium angesehen werden können. Alle Items mit höheren $\mathcal{L}_i$ Werten werden verworfen.

Allerdings sollte man an Hand der $\mathcal{L}_i$ Werte der übriggebliebenen Items überprüfen, ob von diesen das Kontinuum noch genügend gleichmäßig abgedeckt wird, oder ob "leere Stellen" entstanden sind. In diesem Fall müßten für den in Frage kommenden Bereich neue Items konstruiert und getestet werden.

d. Die restlichen 50 % der Items werden nun einem zweiten Ausleseverfahren unterzogen:

Auf Grund der $\mathcal{L}_i$ Werte können wir die Items grob in zwei Gruppen aufteilen, solche die in Bezug auf die getestete

Einstellung positiv und solche, die negativ sind. Jedes dieser Items wird nun mit sechs Antwortkategorien ("stark dafür, dafür, etwas dafür, etwas dagegen, dagegen, stark dagegen") versehen und die Beurteiler gebeten, ihre Einstellung zu den in den Items enthaltenen Behauptungen oder Feststellungen über das Untersuchungsobjekt in Form der Antwortkategorien anzugeben. Die Kategorien werden mit 0,1,2,3,4,5 gewichtet und zwar so, daß immer die positivste Einstellung das größte Gewicht erhält. Also bei den Items der positiven Hälfte die Kategorie "stark dafür", bei denen der negativen Hälfte die Kategorie "stark dagegen".

Jetzt kann jedem Befragten eine "Note" zugeordnet werden, die sich aus der Summe seiner gewichteten Antworten auf alle Items ergibt. Das heißt aber, diejenigen Befragten, die eine stark negative Einstellung zu dem Untersuchungsobjekt haben, werden eine sehr niedrige Note erhalten, umgekehrt alle die positiv zu ihm stehen eine hohe. Nun greifen wir zwei Gruppen von Befragten heraus: etwa die 25% mit der höchsten und die 25% mit der niedrigsten "Note". Damit haben wir zwei Gruppen mit stark entgegengesetzter Einstellung, zwei Gruppen also, die sich auf die beiden Enden des Kontinuum konzentrieren.

Betrachten wir jetzt andererseits die Items. Jedem Item kann ebenfalls (als Durchschnitt der gewichteten Antworthäufigkeiten) eine bestimmte Note zugeordnet werden. Berechnet man nun diese Note für die positive und negative Gruppe resondert, so müßte - vorausgesetzt, alle Items wären perfekt - jedes Item, das von der positiven Gruppe eine hohe Note bekom-

men hat, von der negativen niedrig bewertet worden sein und umgekehrt. Je größer also die Differenz zwischen beiden Noten ist, umso sicherer müßte die Position des Item auf dem Kontinuum sein, je geringer, umso größer ist die Wahrscheinlichkeit, daß das Item fehlerhaft ist. Oder mit anderen Worten, je größer die Differenz, umso größer die Fähigkeit des Item zu diskriminieren und umgekehrt.

Die Differenzen zwischen den Noten, die den Items von der positiven und negativen Gruppe gegeben wurden, könnten also schon als erstes Maß für die Diskriminanzfähigkeit genommen werden. Es ist indessen relativ grob. Genauere Information liefert uns $\mathcal{L}_3$. Betrachten wir in einem Beispiel die ungeachteten Antwortmöglichkeiten für ein Item:

Antwortkategorie	neg. Gruppe	pos. Gruppe
1. stark dafür	2	52
2. dafür	6	44
3. etwas dafür	10	12
4. etwas dagegen	31	6
5. dagegen	38	4
6. stark dagegen	13	2
	100	100

Diese sechs Antwortkategorien können wir nun in zwei Gruppen zusammenfassen, (a,b) und (c,d). Die Trennung zwischen den beiden Gruppen wird nicht immer unbedingt zwischen Kategorie 3 und 4 erfolgen, die einzige Bedingung ist nur, daß die Summe $a + d$ ein Minimum wird. Hätten wir z.B. in der positiven Gruppe eine Verteilung von 38 - 47 - 6 - 5 - 3 - 1, so würden wir die Trennungslinie zwischen Kategorie 2 und 3 ziehen, da

$ln' + 2n + 3p + 4p + 5p + 6p = 23$ kleiner als $ln + 2n + 3n + 4p + 5p + 6p = 27$.

Als Maßzahl für die Diskriminanzfähigkeit bietet sich nun - hauptsächlich wegen seiner Einfachheit - der Kontingenzkoeffizient Pearsons¹⁾ an. Wir können dann $\mathcal{L}_3$ definieren als

$$(5) \quad \mathcal{L}_3 = \frac{bc - ad}{\sqrt{(a+b)(b+d)(a+c)(c+d)}}$$

Theoretisch kann $\mathcal{L}_3$ zwischen +1 und -1 schwanken, wird aber wegen unserer Stellung, daß $a + d$ ein Minimum wird, keinen negativen Wert annehmen.

Der Rechenaufwand bei der Bestimmung von $\mathcal{L}_3$ ist minimal, wenn man entweder den Guilford'schen Monograph oder die Tabellen von Jurgensen verwendet.²⁾

Bei der Auswahl der endgültigen Items wird man nun am besten so vorgehen: die noch vorhandenen Items werden nach ihren $\mathcal{L}_3$ Werten geordnet und auf dem Kontinuum in so viele Gruppen zusammengefaßt, wie endgültige Items beabsichtigt sind. Aus jeder Gruppe wird dann das Item mit dem höchsten $\mathcal{L}_3$ Wert ausgewählt. Diese restlichen Items stellen dann den Satz an "geeignetesten" Items für die spätere Skalierung dar.

4. Zusammenfassung

a. Das Ziel dieses Abschnitts ist es, aus der Gesamtheit der möglichen Items diejenigen herauszufinden, die sich am

1) Kendall, W.G.: The Advanced Theory of Statistics, Vol I, London, 1952.

2) Guilford, J.P.: The phi coefficient and chi square as indices of item validity. Psychometrika (6), 1941, S.11-19.
Jurgensen, C.E.: Table for determining phi coefficients, Psychometrika (2), 1947, S.17-29.

besten für die Skalierung des Untersuchungsobjekts eignen. Diese Items müssen sich, kurz gesagt, möglichst gleichmäßig über den relevanten Teil des Kontinuum verteilen und sie müssen gut diskriminieren. Mit anderen Worten, sie sollen nicht doppeldeutig, unscharf, ohne direkten Bezug auf das Kontinuum usw. sein. Eine intuitive oder zufallsauswahl kann diese Kriterien offensichtlich nicht erfüllen. Wir haben deshalb drei Maße aufgestellt (die alle der elementaren Statistik entstammen!), die eine objektive Auswahl nach den oben genannten Kriterien ermöglichen sollen.

Über $\mathcal{L}_1$ können wir abschätzen, wie sich die vorgeschlagenen Items auf dem Kontinuum verteilen. $\mathcal{L}_2$ bietet ein erstes "Sieb", um fehlerhafte Items auszuschneiden - besonders diejenigen, deren Position auf dem Kontinuum unklar ist, werden entfernt. Die Berechnung von $\mathcal{L}_3$ schließlich läßt eine Auslese nach dem Grad der Diskriminanzfähigkeit der Items zu. $\mathcal{L}_1$ und $\mathcal{L}_3$ bieten natürlich nur eine Art von Fausregel für das Verwerfen von Items. "Kritische Werte" etwa, ober- oder unterhalb derer ein Item absolut abzulehnen ist, können nicht angegeben werden. Indessen sind $\mathcal{L}_1$ und $\mathcal{L}_3$ eine objektivere Basis für die Auswahl der geeignetsten Items als intuitive Kriterien. Damit sind sie diesen überlegen.

b. Die erste Stufe unseres Modells haben wir damit erklommen. Wir besitzen eine bestimmte Zahl plausibler Items, von denen wir objektiv wissen, daß sie die besten der vorgeschlagenen sind und die wir rangmäßig auf dem Kontinuum anordnen können. Sie bilden allerdings noch keinen metrischen

Maßstab. Die $\mathcal{L}_1$ Werte waren für uns nur Hilfsgrößen, sie liefern keine echte metrische Information.

Die zweite Stufe des Modells, die Auswahl der Befragten, können wir überspringen. Das Problem der Auswahl einer repräsentativen Teilgesamtheit wird in der Stichprobentheorie ausgiebig behandelt, es liegt theoretisch auf einem anderen Gebiet als die Skalierungstechnik. Wir dürfen daher hier auf die umfangreiche Literatur verweisen.¹⁾

-
- 1) Kellerer, H.: Theorie und Technik des Stichprobenverfahrens, München, 1953.
Hobart, D.M.: Praxis der Marktforschung, deutsche Ausgabe von C.Hundhausen, Essen, 1952.
König, R., a.a.O., 1952.
Deming, W.E.: Some Theory of Sampling, New York, 1950.
Deming, W.E.: Sample Decision in Business Research, New York, 1960.
Yates, F.: Sampling Methods for Censuses and Surveys, London, 1949.
Bradford, E.S.: Marketing Research: How to Analyse Products, Markets and Methods of Distribution. New York, 1951.
Parten, M.B.: Surveys, Polls and Samples. New York, 1950.
Blankenship, A.B. (ed.): How to Conduct Consumer and Opinion Research. The Sampling Survey in Operation, New York, London, 1946.
Tagliacarne, G.: Tecnica e pratica delle ricerche di mercato, Mailand, 1951.
Horn, G.H.: Meinungsforschung, AStA, München, 1952.
Kellerer, H.: Marktforschung - ein Anwendungsgebiet des Stichprobenverfahrens. MMS, München, 1952.
Neyman, J.: On two different aspects of the representative method: the method of stratified sampling and the method of purposive selection, JRS., 1934.

B. Die Konstruktion der Skala

Vorbemerkung

a. Wir sind nun im Besitz einer Reihe plausibler Items einerseits, die uns als Maßstab für die zu untersuchende Eigenschaft dienen sollen, und einer repräsentativen Teilgesamtheit der zu befragenden Träger dieser Eigenschaft andererseits. Jetzt können wir mit der eigentlichen Messung beginnen. Da unsere Items nicht auf einer metrischen Skala gegeben und die Modelle, die wir in dem vorliegenden Kapitel besprechen wollen, streng deterministisch sind, werden wir nur in der Lage sein, eine topologische Skala zu konstruieren. Das heißt, wir werden nur eine rangmäßige Anordnung der Befragten in Bezug auf die betreffende Eigenschaft erzielen können.

Unsere Aufgabe ist es nun hier Methoden aufzuzeigen, über die wir von den Items zu der (topologischen) Messung der Einstellung selbst kommen. Das Problem erscheint zuerst trivial. Man brauche nur, so sollte man meinen, die rangmäßig geordneten Items den Befragten vorzulegen und aus den Antwortäußerungen auf die Einstellungsverteilung zu schließen. Das ist auch ohne Zweifel das Prinzip der besprochenen Methoden - indessen, allein genügt es nicht. Zwei Fragen müssen zusätzlich geklärt werden: es muß eine Testmethode gefunden werden, die es erlaubt, die zugrundeliegende Hypothese der Eindimensionalität zu testen und es muß - wenn auch die Messung nur topologisch ist - zumindest der genaue O-Punkt festgestellt werden, also der Punkt, der angibt, wieviel der Befragten nun effektiv "dafür" und wieviel "dagegen" sind.

Durch die Kriterien $\mathcal{L}_1$, $\mathcal{L}_2$ und $\mathcal{L}_3$, konnten wir zwar ungeeignete Items ausscheiden, einen Test auf Eindimensionalität bieten sie indessen nicht. Da den Beurteilern bei der Berechnung von $\mathcal{L}_1$ und $\mathcal{L}_2$ vorgeschrieben ist, die Items in Bezug auf eine bestimmte Eigenschaft zu ordnen, wird den Daten Transitivität - die Voraussetzung eines eindimensionalen Kontinuum - quasi aufgezwungen. Das heißt, die Möglichkeit weiterer Dimensionen, die für Übereinstimmung oder nichtÜbereinstimmung der Befragten mit den einzelnen Items entscheidend sein können, wird von vornherein nicht berücksichtigt. Die so gewonnene Anordnung der Items muß also nicht invariant sein. Wir brauchen daher einen zusätzlichen Test auf das Vornandensein nur einer Dimension.¹⁾ Wir werden ihn im folgenden erhalten.

Zweitens, der O-Punkt: die Einteilung zwischen positiven und negativen Items bei der Berechnung von $\mathcal{L}_3$, war nur eine - für die Aufgabe damals genügende - Schätzung oder beinahe willkürliche Festsetzung des O-Punktes. Gerade auf seine genaue Bestimmung kommt es uns aber an. Das Hauptinteresse bei einer Meinungsumfrage liegt zuerst bei der Frage: wer ist nun eigentlich "dafür" und wer "dagegen", das Ziel unserer Arbeit ist es ja u.a. auch hier zu genaueren Ergebnissen zu kommen, als man durch direkte oder intuitive Fragen erhalten kann. Das Problem des O-Punktes muß uns also beschäftigen. Wir

1) Das Problem der Dimensionalität wird vielleicht am besten an einem Beispiel klar. Mehrere Dimensionen einer Einstellung haben wir - salopp formuliert - dann, wenn der Befragte geneigt ist, auf bestimmte Items mit "ja, aber..." zu antworten. Wollten wir etwa die "Einstellung zur UNO" messen und würden ein Item formulieren: "Die UNO ist eine große Errungenschaft unserer Zeit", so würden vielleicht manche Befragte gern darauf antworten: "Ja, aber Kriege kann sie doch nicht verhindern." Offensichtlich besitzt die "Einstellung zur UNO" mehrere Dimensionen. Man müßte sie unterteilen, zum Beispiel in "Einstellung zur UNO als Friedensvermittlerin" usw.

werden sehen, daß eine sogenannte "intensity analysis" zu ihm führen kann.

b. Soviel kann zu den Zielen der folgenden Abschnitte gesagt werden. Die Methoden, derer wir uns bei ihrer Erreichung bedienen werden, können wir in zwei Gruppen einteilen: Einmal, die naheliegende Methode, die Items den Befragten einfach vorzulegen und sie zu bitten anzugeben, mit welchen sie übereinstimmen und mit welchen nicht. Für jedes Item werden sich mithin zwei Kategorien von Befragten ergeben: diejenigen, die dafür und diejenigen, die dagegen sind. Methoden, die auf diesem Grundgedanken aufbauen, wollen wir "kategorische Methoden" nennen. Im zweiten Fall ist die Aufgabe der Befragten eine andere. Sie sollen jetzt nicht einfach mit "ja" oder "nein" auf jedes Item antworten, sondern die einzelnen Items nach ihrer subjektiven Einstellungsskala ordnen. Also zuerst dasjenige, das ihnen am sympathischsten ist, dann das zweisymphatische usw. Diese Aufgabe ist natürlich für den Befragten schwieriger als die erste, andererseits können wir dadurch beinahe schon zu einer quasi-metrischen Skala gelangen. Wir wollen dieses Vorgehen mit "komparative Methoden" bezeichnen.

I. Kategorische Methoden.

1. Der Grundgedanke

Das sehr einfache Modell der kategorischen Messung hatten wir einleitend schon einmal erwähnt. Hätten wir z.B.

vier perfekte, monoton-dichotome Items, so würde sich - bei vorhandener Eindimensionalität - folgendes "Skalogramm" ergeben.¹⁾

Schaubild 6

	Skala I				Skala II			
	1	2	3	4	1	2	3	4
I	X	X	X	X				
II		X	X	X	X			
III			X	X	X	X		
IV				X	X	X	X	
V					X	X	X	X

Das heißt, im Falle der Idealskala werden bei n Items, von den 2^n möglichen Antwortkombinationen nur $n+1$ erscheinen. Das Nichterscheinen atypischer Kombinationen ist die Grundlage der Messung: die Zeilen und Spalten des Skalogram werden solange geordnet bis sie das obige typische Bild ergeben. Damit wären wir praktisch fertig, Befragte und Items können rangmäßig auf dem Kontinuum angeordnet werden.

1) Zu dem hier vorgeführten Modell siehe vor allem:
Guttman, L.: Kapitel 2,3,6,8 und 9 in: Stouffer, A.A. et al.: Measurement and Prediction, Princeton, 1950.
Guttman, L.: a.a.O., 1941 (hier findet sich die einzige klare Darstellung seines Modells).
Suchman, E.A.: Kapitel 4,5 und 7 in: S.A. Stouffer et al., a.a.O., 1950.
Guttman, L.: The principal components of scalable attitudes, in P.F. Lazarsfeld: Mathematical Thinking in the Social Sciences, Glencoe, Ill., 1954, S.216-257.
Edwards, A.L.: a.a.O., 1957, S.172-201.
Torgerson, W.S.: a.a.O., 1958, S.293-345.

In der Praxis werden indessen atypische Antwortkombinationen erscheinen. Es wird nie möglich sein, Zeilen und Spalten des Skalogramms so zu ordnen, daß sie exakt der Idealskala entsprechen. Damit ergeben sich zwei Probleme:

Einmal ein praktisches: wie ordne ich die verschiedenen Antwortkombinationen an, so daß sie so gut als möglich der Idealskala entsprechen?

Zweitens: wie kann ich das Ausmaß feststellen, mit dem die einzelnen Daten auf das Modell passen, sind die Abweichungen zufällig oder systematisch, ist mithin die zugrundeliegende Hypothese der Eindimensionalität richtig oder nicht?

2. Auswertungsmethoden

a. Methoden für eine bestmögliche Anordnung der Antwortkombinationen. Das Hauptproblem ist hier Items und Befragte so zu ordnen, daß das Idealmodell so weitgehend wie möglich angenähert wird. Eine Vielzahl von Methoden ist vorgeschlagen worden. Die einfachste ist die Suchmans.¹⁾ Sie besteht in der recht ingeniösen Erfindung des "Skalogramm-Brett".

Das Skalogramm-Brett besteht aus zwei Tafeln, in die herausziehbare Schieber eingelassen sind, die man in einer Tafel horizontal, in der anderen vertikal herausziehen und austauschen kann.

1) Eine eingehende Darstellung dieser Technik, mit Abbildungen und Beispielen, findet sich bei Suchman, E.A.: a.a.O., 1950, Kapitel 4.

Die vertikalen Schieber entsprechen den Items, die horizontalen den Befragten. Auf den Schiebern sind kleine Vertiefungen, entsprechend den Zellen der Antwortmatrix. Für jede positive Antwort werden kleine Kügelchen in die betreffende Vertiefung gelegt und zwar zuerst in die des Bretts mit den vertikalen Schiebern. Eine Umordnung dieses Schiebers entspricht nun einer Umordnung der Items. Andererseits entspricht eine Umordnung der horizontalen Schieber einer Umordnung der Befragten.

Die Bretter sind so konstruiert, daß man die Kügelchen leicht in dem einen zum andern Übertragen kann, indem

Skalogramm-Brett



man einfach beide Bretter übereinanderlegt und das Ganze umdreht. Auf diese Art kann man solange beliebig zwischen einer Umordnung der Items und einer Umordnung der Befragten hin und her wechseln, bis die beste Annäherung an das Idealmodell erreicht ist. Eine ungefähre Vorstellung von der Rangordnung der Items und der Befragten ist natürlich als Ausgangsbasis notwendig.

Mit dieser Technik kann man selbstverständlich nur Daten aus einer kleinen Befragung verarbeiten. Das von Suchman in der genannten Veröffentlichung abgebildete Skalogramm-Brett läßt sich z.B. nur bis zu einer Stichprobe von 100 Befragten benutzen.

b. Eine zweite Methode der Anordnung der Antwortkombinationen ist die von Guttman vorgeschlagene "Cornell-Technik".¹⁾ Sie stellt im Grunde nur eine Übertragung des Skalogramm-Brett auf Papier dar. Eine Tabelle wird aufgestellt, in der jede Spalte einem Item und jede Reihe einem Befragten entspricht. Die Befragten bekommen nun zuerst - auf Grund der Fragebögen - "Noten", die sich aus der Zahl, der von ihnen positiv beantworteten Items berechnen. Beginnend mit dem höchstbenoteten Befragten werden die Antworten der Einzelnen auf der Tabelle eingetragen (ja=1, nein=0).

Schaubild 7

Be- fragte	Items				Noten	Summe
	1	2	3	4		
1	1	1	1	1	4	4
2	1	1	1	1	4	4
3	1	0	1	1	3	4
4	1	0	1	1	3	4
5	0	1	1	1	3	4
6	0	1	1	1	3	4
7	0	1	1	1	3	4
8	0	1	1	1	3	4
9	0	0	1	1	2	2
10	1	0	0	1	2	2
11	0	0	1	1	2	2
12	0	0	1	1	2	2
13	1	0	0	1	2	2
14	0	1	0	1	2	2
15	1	0	0	1	2	2
16	0	0	0	1	1	1
17	0	0	1	1	2	2
18	0	1	0	1	2	2
19	0	0	0	1	1	1
20	0	0	0	0	0	0
Σ	6	8	12	16	42	72

1) Guttman, L.: a.a.O., 1950, S.6-90.

Im Idealfall müßte man jetzt aus der Rangziffer der Befragten seine Antwortkombination reproduzieren können. Das heißt, bei vier Items (Schaubild 6) müßte ein Befragter mit der Rangziffer 2, Item 3 und 4 bejaht, Item 1 und 2 verneint haben (--++). Ist das empirisch nicht der Fall, sind Antwortkombinationen wie (++++), oder (-++-) vorhanden, so müssen die Befragten mit atypischen Antwortkombinationen einer bestimmten Gruppe zugeordnet werden. Dies geschieht am besten dadurch, daß man die Rangziffern so neu verteilt, daß die Gesamtzahl der Abweichungen ein Minimum wird (siehe Schaubild 7). Als zusätzliche Bedingung gilt, daß für kein Item mehr Abweichungen als Nichtabweichungen vorhanden sind.

Ähnliche, ebenfalls auf Tabellierung aufbauende Techniken wurden von Goodenough¹⁾ und Marder²⁾ entwickelt. Abwandlungen dieser Methode, die die Anwendung von IBM Rechenautomaten erlauben, werden von Noland³⁾, Ford⁴⁾, Kahn und Bodine⁵⁾ sowie Toby⁶⁾ beschrieben.

c. Guttman⁷⁾ schließlich bietet noch eine Lösung des Problems mit Hilfe der Methode der kleinsten Quadrate. Sie

-
- 1) Goodenough, W.H.: A technique for scale analysis. Educ. Psych.Msmt., 1944 (4), S.179-190, nach Edwards, A.L., a.a.O., 1957.
 - 2) Marder, E.: Linear segments: a technique for scalogram analysis. Publ.Op.Quart., 1952 (16), S.417-431.
 - 3) Noland, E.W.: Worker attitude and industrial absenteeism: a statistical appraisal. Am.Soc.Rev., 1955 (10), S.503-510.
 - 4) Ford, R.N.: A rapid scoring procedure for scaling attitude questions. Publ.Op.Quart., 1950 (14), S.507-532.
 - 5) Kahn, L.A. und A.J. Bodine: Guttman scale analysis by means of IBM equipment. Educ.Psych.Msmt., 1951 (11), S.298-340, nach Torgerson, W.S.: a.a.O., 1958.
 - 6) In Riley, W. et.al.: Sociological Studies in Scale Analysis, New Brunswick, N.J., 1954.
 - 7) Guttman, L.: a.a.O., 1950.

besticht zwar durch ihre mathematische Eleganz und liefert unzweifelhaft eine "beste" Lösung, erfordert aber weit mehr rechnerischen Aufwand als die bisher präsentierten, primitiven "cut and try" Methoden. Sie wurde praktisch bisher noch kaum angewandt, wir wollen es daher bei ihrer Erwähnung belassen.

Kehren wir kurz zu unseren Ausgangsüberlegungen zurück. Legt man die ausgewählten Items den Befragten vor, so muß sich im Idealfall das Parallelogramm-Bild des Skalogramm ergeben (Schaubild 6). Dadurch werden Items und Befragte rangmäßig geordnet, die Rangordnung ist invariant, die Häufigkeiten auf den einzelnen Rangziffern sind bekannt, die topologische Messung ist damit abgeschlossen. Da indessen atypische Antwortkombinationen in der Praxis auftauchen, entstand die Aufgabe, mit dem vorhandenen Datenmaterial das Idealbild so weitgehend als möglich anzunähern und zweitens, ein Maß für den Grad dieser Annäherung und damit - wie wir sehen werden - einen Test auf die zugrundeliegende Hypothese der Eindimensionalität zu finden. Die erste Aufgabe haben wir - wenn auch mit relativ rohen Mitteln - gelöst. Nun zur zweiten.

3. Der Test auf Eindimensionalität

a. Unsere Argumentation war: wenn die untersuchte Eigenschaft ein eindimensionales Kontinuum besitzt, dieses durch die Items scharf abgeteilt wird und sich die Befragten andererseits "ihrer" Einstellung klar bewußt sind, so muß die rangmäßige Anordnung von Items und Befragten auf dem Kontinuum invariant sein. Das heißt aber, daß wir - erinnern wir uns an das Bild

des Skalogramm - aus der Rangziffer jedes Befragten seine Antwortkombination genau reproduzieren können.

Damit können aber auch die Abweichungen, die zwischen empirischen Material und Modell bestehen, als Maß für die Reproduzierbarkeit und damit als ein, wenn auch grober, Test auf die Hypothese der Eindimensionalität benutzt werden.

Es ließe sich folgende Maßzahl konstruieren (R = Reproduzierbarkeit):

$$(6) \quad R = 1 - \frac{A}{Nn}$$

wobei A Zahl der Abweichungen

N Zahl der Items

n Zahl der Befragten

Ein R von 0,96 würde dabei etwa bedeuten, daß wir aus den Rangziffern der Befragten die ursprünglichen Antwortkombinationen zu 96 % reproduzieren können.

Ebenso läßt sich die Reproduzierbarkeit für jedes einzelne Item berechnen. Sie ist

$$(7) \quad R_j = 1 - \frac{A_j}{n}$$

R kann mithin als Kriterium dafür betrachtet werden, wie gut sich die empirischen Daten der Idealskala annähern, das heißt aber auch, inwieweit die Hypothese eines eindimensionalen Kontinuum zutrifft. Als Grenze für Annahme oder Verwerfen dieser Hypothese kann man - willkürlich - ein R von 0,85 oder 0,90 festsetzen.

Das Kriterium R allein reicht allerdings nicht aus, um die Skalierbarkeit einer Einstellung ganz zu beurteilen. Zu-

sätzliche Kriterien sind notwendig:

a) Da

$$(8) \quad R = \bar{R}_I$$

R_I jedoch nie geringer sein kann als die relative Be-
setzung der am stärksten vertretenen Antwortkategorie,
werden extreme Items, auf die etwa mehr als 80% der Be-
fragten mit "ja" oder "nein" antworten, R rein formal in
die Höhe drücken, obwohl diese Items theoretisch völlig
unabhängig sein können. Man wird also zusätzlich darauf
achten müssen, daß keine Antwortkategorie zu mehr als
80% (oder 90%, je nach dem angestrebten Grad an Genau-
igkeit) besetzt ist - oder aber - wenn man gerade auch
extreme Einstellungen erfassen will, diese bei der Be-
rechnung von R nicht berücksichtigen.

Zweitens sollte man prüfen, ob sich die Abweichungen zu-
fällig verteilen oder nicht. Das heißt, es sollten keine
größeren Befragtengruppen existieren, die die gleichen
atypischen Antwortkombinationen zeigen.¹⁾

c) Schließlich sollte auch R_I für jedes einzelne Item
nicht niedriger als 0,85 sein.

Alle diese Kriterien geben natürlich nur eine Art
Fausregel^t. Mehr läßt sich jedoch nicht erreichen, da das Mo-
dell in seinem Aufbau streng deterministisch ist. Das heißt:
im Modell selbst ist auf keinerlei Art und Weise eine Theorie

1) Umgekehrt ist es durchaus möglich, daß sich für eine Ein-
stellung ein unter 0,85 liegendes R ergibt, alle Abwei-
chungen aber zufällig sind. Es ergibt sich dann eine sog.
Quasi-Skala, die durchaus noch verwendet werden kann.

der Abweichungen eingebaut. Statistische Testverfahren können daher nicht angewendet werden.

b. Die eben vorgeführte Berechnung von R erfordert allerdings eine aufwendige Manipulation des Datematerials. Das bedeutet aber, daß dieses Verfahren nur auf Befragungen mit relativ wenig Items und Befragten beschränkt bleiben muß. Wenn man nun weniger an der genauen Rangordnung der Befragten interessiert ist, sondern mehr an der Frage, ob die untersuchte Einstellung überhaupt in der geforderten Form vorhanden ist, so kann man eine durchaus exakte Methode der Berechnung von R anwenden, die vom Urmaterial (den Fragebogen) selbst ausgeht und keine tabellarische oder andere Umarrangierung von Items und Befragten erfordert.¹⁾

Die Gesamtanzahl der Abweichungen in einer Antwortkombination kann als Summe der Abweichungen ersten, zweiten und dritten Grades definiert werden. Dabei versteht man unter Abweichungen verschiedenen Grades folgendes: Angenommen wir hatten eine Antwortkombination von der Form (+-----+). Die dieser Kombination ähnlichste Idealkombination wäre (-----+). Abweichungen 1. Grades entsprächen nun den Itempaaren, die in verkehrter Reihenfolge beantwortet wurden, also (+-) statt (-+). Zwei solche Paare sind hier vorhanden, wir haben also zwei Abweichungen 1. Grades. Eliminiert man diese Itempaare, so ergibt sich (----+---+). Hier ist noch immer

1) Green, B.F.Jr.: A method of scalogram analysis using summary statistics. Psychometrika, 1956 (21), S.79-89.
Torgerson, W.S.: a.a.O., 1958, S.325 f.

ein (+-) Paar vorhanden. Dies nun entspricht einem Fehler 1. Grades. Wird auch diese Abweichung eliminiert, so ist mit (---+++) eine perfekte Kombination erreicht. Fehler 3. Grades sind nicht mehr vorhanden. Empirisch hat sich gezeigt, daß Fehler 3. Grades fast nie auftauchen, man kann sie also ohne große Gefahr ignorieren. Die Summe der Abweichungen ist dann

$$(9) \quad A = A_1 + A_2$$

Abweichungen 1. Grades erscheinen in den Fragebögen als (+-), Abweichungen 2. Grades als (+-+). Das heißt

$$(10) \quad A_1 = \sum_{i=1}^{n-1} N_{i+1, \bar{i}} \quad (i+1) > i$$

wobei $N_{i+1, \bar{i}}$ die Anzahl der Befragten bezeichnet, die positiv auf Item $i+1$ und negativ auf Item i geantwortet haben. Item $i+1$ hat dabei eine höhere Rangziffer als Item i .

Ferner ist

$$(11) \quad A_2 = \sum_{i=2}^{n-1} N_{i+2, i+1, \bar{i}, \bar{i}-1}$$

die Summe der Abweichungen 2. Grades und daher

$$(12) \quad R = 1 - \frac{1}{N \cdot n} \sum_{i=1}^{n-1} N_{i+1, \bar{i}} - \frac{1}{N \cdot n} \sum_{i=2}^{n-1} N_{i+2, i+1, \bar{i}, \bar{i}-1}$$

der Reproduzierbarkeitsfaktor.

Das heißt, wir müssen bei dieser Methode der Berechnung von R praktisch nur zuerst auf Grund der Antworthäufigkeiten die Rangordnung der Items feststellen, dann aus den Fragebögen A_1 und A_2 abzählen und R nach (12) berechnen. Es braucht also keine der Tabellen, die oben beschrieben wurden, konstruiert zu werden.

In (6) und (12) haben wir nun zwei Möglichkeiten, R zu berechnen. Damit ist wenigstens ein grober Test auf die Richtigkeit der Hypothese der Eindimensionalität gegeben. Somit ist auch dieser Teil des Problems "rangmäßige Skalierung einer Einstellung" gelöst.

Eine letzte Frage bleibt noch zu klären. Gerade bei der Erforschung von Meinungen und Einstellungen kommt es zuerst immer darauf an zu bestimmen, wieviele der Befragten nun eigentlich "für" und wieviele "gegen" das Untersuchungsobjekt sind. Es ist ja gerade eines der Ziele der Skalierungstechnik hier genauere und unverzerrtere Angaben zu liefern, als mit direkten oder intuitiven Fragen erzielt werden können. Die rangmäßige Ordnung der Items und Befragten, die wir bis jetzt erzielt haben, läßt aber noch keinen Schluß auf diesen "O-Punkt" zu. Auch die Einteilung, die wir bei der Auswahl der Items für die Berechnung von $\mathcal{L}$, getroffen haben, war nur eine Hilfsgröße und genügt nicht zur scharfen Trennung des Pro und Kontra. Wir werden also noch eine Methode finden müssen, die hier exakte Schlüsse zuläßt.

4. Die Bestimmung des O-Punkts

a. Es erscheint plausibel, die Intensität, mit der sich die Befragten für oder gegen ein Item entschieden haben, zur Bestimmung des O-Punktes zu benutzen. Befragte, die eine extreme Einstellung, also eine Position an den beiden Enden des Kontinuum besitzen, werden im allgemeinen diese Einstellung mit größerer Emphase vertreten als diejenigen, die sich mehr

in der Mitte, dem neutralen Teil des Kontinuum befinden. Es ist also vernünftig den Punkt, an dem die Befragten am indifferentesten gegenüber dem Untersuchungsobjekt erscheinen, als O-Punkt zu definieren.

Schaubild 7

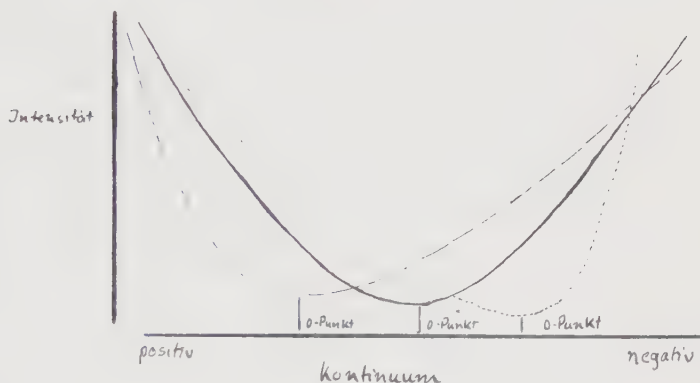


Schaubild 7 zeigt den Verlauf einiger solcher Intensitätskurven. Projiziert man ihr Minimum auf die Skala des Kontinuum, so ergibt sich der O-Punkt.

b. Um nun den O-Punkt auf diese Art und Weise zu berechnen, brauchen wir eine Methode für die Messung der Intensität der Antworten auf die einzelnen Items. Erinnern wir uns an die Berechnung von $\mathcal{L}$. Ähnlich können wir auch hier vorgehen. Jedes einzelne Item wird nicht nur mit zwei Antwortkategorien "ja" und "nein" versehen, sondern eine zweite Frage (die Intensitätsfrage), etwa der Form "Wie sicher sind Sie

sich Ihrer Antwort?" oder so ähnlich wird daneben geschaltet, auf die der Befragte in der bekannten Weise mit "stark dafür, dafür usw." antworten kann.¹⁾ Dabei ist die genaue Formulierung dieser Zusatzfragen nicht allzu wichtig, die Ergebnisse werden immer ungefähr gleich sein.

Beispiel eines Fragebogens für 9 Items mit 5 Intensitätsfragen

Items \ Intensitätsfragen	stark dagegen	dagegen	neutral	dafür	stark dafür
1.	+				
2.	+				
3.	+				
4.		+			
5.		+			
6.			+		
7.			+		
8.				+	
9.				+	

1) Suchman, E.A.: a.a.O., 1950, S.217 ff.

Die Antworten auf diese "Intensitätsfragen" werden nun einer ganz normalen Skalogramm-Analyse unterzogen. Obwohl R für sie meist zu niedrig sein wird, werden sich die Abweichungen genügend unsystematisch verteilen. Damit ist die erhaltene Skala als Quasi-Skala qualifiziert (siehe Gußnote, Seite 50). Auf diese Art und Weise erhalten wir nun zusätzlich zu den Einstellungsrangziffern der Befragten, sogenannte Intensitäts-Rangziffern. Jetzt kann der O-Punkt auf verschiedene Methoden bestimmt werden. Am einfachsten wird es sein, den Median der Einstellungsrangziffern derjenigen Befragten zu berechnen, die die niedrigsten Intensitäts-Rangziffern besitzen. Voraussetzung ist dafür natürlich eine relativ große Stichprobe (mindestens über 100 Befragte).

c. Der so erhaltene O-Punkt wird von Verzerrungen in den Fragen unabhrängig sein. Dies ist eine äußerst wichtige Eigenschaft. Sie erweist sich besonders dann als nützlich, wenn man die Ergebnisse verschiedener Befragungen zu einer Einstellung, die zu verschiedenen Zeitpunkten mit verschiedenen Items stattgefunden haben, miteinander vergleichen will. Suchman, der sich derartiger Methoden häufig bedient hat,¹⁾ machte folgendes Experiment: von zwei verschiedenen 12-Item-Skalen nahm er je die sechs positivsten und die sechs negativsten Items und legte sie als gesonderte Skala einer bestimmten Einstellung zusammen mit Intensitätsfragen verschiedenen Befragtengruppen vor. Obwohl sich in beiden Fällen eine sehr unterschiedliche rang-

1) Suchman, E.A.: a.a.O., 1950, S.266-275.

mäßige Verteilung ergab, teilte der O-Punkt die Befragten beinahe im gleichen Verhältnis. Das heißt, die O-Punkte erwiesen sich trotz der extremen Verzerrung der Skalen als konstant.

5. Zusammenfassung

Ziel des vorliegenden Abschnitts ist es, kategorische Methoden der Konstruktion einer topologischen Skala für eine bestimmte Einstellung zu entwickeln. Wir haben jetzt gezeigt, wie wir von dem Satz ausgewählter Items und der repräsentativen Teilgesamtheit der zu untersuchenden Befragten, die wir uns im vorigen Abschnitt erarbeitet hatten, zu der topologischen Messung der Verteilung der Einstellung zu kommen. Die Rangziffern sowohl der Items, wie der Befragten sind damit bekannt. Darüber hinaus haben wir eine Art Test gezeigt, der es erlaubt, die zugrundeliegende Hypothese der Eindimensionalität zu prüfen. Bis dahin war noch nicht sicher, ob die Einstellung ein mehrdimensionales Kontinuum besitzt, das heißt, sich eine Reihe psychischer Variablen in ihr vereinigen. Schließlich haben wir noch eine Methode zur Bestimmung des sog. O-Punkts dargestellt, mittels derer wir exakt und vergleichbar feststellen können, welcher Teil der Befragten wirklich "dafür" und welcher "dagegen" ist.

Dieser "kategorischer" Skalierung der Befragten steht die "komparative" gegenüber. Ohne Zweifel könnten wir aus den Items noch mehr Information herausholen, wenn wir die Aufgabe der Befragten so verändern, daß diese nicht nur Übereinstim-

mung und Nichtübereinstimmung mit jedem Item bekanntgeben, sondern diese praktisch auf einer Präferenzskala ordnen, sie also miteinander vergleichen. Diese "komparative" Methode ist nicht nur theoretisch äußerst interessant, sie erlaubt vielmehr auch die Skalierung anderer Stimuli als Items. So können auf diese Art und Weise z.B. Waschmittel oder verschiedene Sorten Speiseeis skaliert werden. Die Methode ist also gerade für die Marktforschung wichtig. Doch dazu im nächsten Abschnitt.

II. Komparative Modelle

1. Grundgedanke

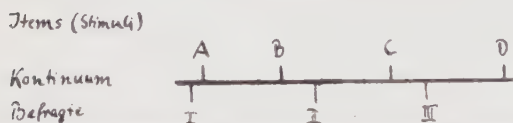
a. Erinnern wir uns an die einleitend gegebene schematische Darstellung des Grundgedankens eindimensionaler Skalierungsmodelle (Schaubild 2., Befragte und Items¹⁾) - oder, wie wir im folgenden allgemeiner sagen wollen, Stimuli - können als Punkte auf dem zugrundeliegenden Kontinuum dargestellt werden. Dabei hat die Entfernung jedes Stimulus von einem Befragten eine direkte Beziehung zu dem Ausmaß, mit dem der einzelne Befragte mit ihm übereinstimmt oder nicht. Ein Stimulus, der sich am gleichen Platz befindet wie der Befragte wird von ihm voll und ganz bejaht werden - mit steigender Entfernung sinkt die Präferenz für die entsprechenden Stimuli. Würden wir mit s_A , s_B und s_C , die auf einer metrischen Skala gegebenen Skalenwerte der Stimuli bezeichnen, dann bedeutet die Präferenz: Stimulus B wird Stimulus C vorgezogen, daß

1) Items bedeutet hier nicht mehr monoton dichotyme, sondern Punktitems.

$$(13) \quad |s_B - s_A| < |s_C - s_A|$$

Die komparativen Methoden bauen auf diesem Grundgedanken auf. Er bedeutet, daß die in den Präferenzen gegebene subjektive Rangordnung der Stimuli für einen Befragten, der Rangordnung der Entfernungen der Stimuli von der Position des Befragten auf dem Kontinuum entspricht. Jeder Befragte hat nun seine eigene Präferenzskala:

Schaubild 8



Wäre die Situation, wie in Schaubild 8 gegeben, so würde der Befragte I die Präferenzsskala (ABCD), II dagegen (BCAD) und III (CDBA) besitzen.

Bei einer empirischen Befragung ist uns allerdings die Position der Items und Befragten auf dem Kontinuum unbekannt. Bekannt sind hingegen die Präferenzskalen der einzelnen Befragten, die ja gebeten wurden, die Stimuli nach dem Grad ihrer Übereinstimmung mit der eigenen Position zu ordnen. Das Problem ist mithin, aus der Information, die wir über die einzelnen subjektiven Präferenzskalen besitzen, die Einstellungsskala zu rekonstruieren.

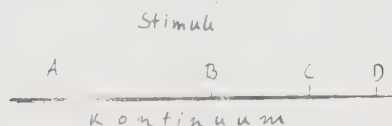
Die Lösung dieses Problems wollen wir im folgenden besprechen. Zuerst für den Fall totaler Präferenzskalen, zweitens für den partieller Präferenzskalen. Im ersten Fall haben die Befragten sämtliche Stimuli nach ihrer Präferenz geordnet, im zweiten nur diejenigen drei oder vier, die ihrer Position

am nächsten sind. Für beide Fälle wird sich ein etwas verschiedenes Vorgehen ergeben.¹⁾

2. Nehmen wir an, N Befragte hatten je n Stimuli nach ihren subjektiven Präferenzen geordnet.²⁾ Es kann sich bei den Stimuli um Einstellungsitems, Speiseeisproben verschiedener Süße, Wahlkandidaten verschiedenen politischen Profils usw. handeln,

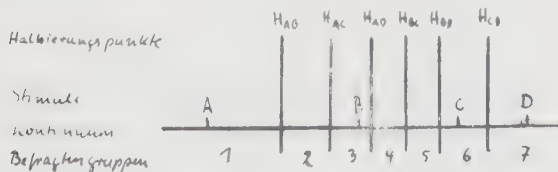
deren relative (uns noch unbekannte) Position dieser Stimuli auf dem Kontinuum wäre wie in Schnaubild 9 ($n = 4$):

Schnaubild 9



Der Halbierungspunkt der Entfernung zwischen zwei Stimuli teilt nur die Befragten in jeweils zwei Gruppen. Diejenigen, die sich auf der einen Seite des Halbierungspunkts befinden, werden die zwei Stimuli in einer bestimmten Reihenfolge ordnen, diejenigen, die sich auf der anderen Seite befinden, genau umgekehrt.

Schnaubild 10



- 1) Coombs, C.H.: a.a.O., 1955.
Coombs, C.H.: Application of the Unfolding Technique to the Study of Stimulus Scales, Manuskript, 1953.
Coombs, C.H.: The Measurement of Social Utility, Manuskript, 1953.
- 2) Das heißt natürlich, daß Transitivität und Konsistenz der Beurteilungen der Befragten vorausgesetzt werden.

In Schaubild 10 werden z.B. alle Befragten aus Gruppe 1, die die Stimuli A und B in der Reihenfolge AB, alle aus 2 in der Reihenfolge BA ordnen. Der Halbierungspunkt zweier Stimuli kann also als Grenze zwischen zwei Befragtengruppen dienen. Bei n Stimuli existieren $1/2 n (n-1)$ Halbierungspunkte und damit $1/2 n (n-1) + 1$ geordnete Befragtengruppen. In unserem Beispiel werden sich, keine Abweichungen des empirischen Materials vom Modell vorausgesetzt, sieben Befragtengruppen mit folgenden subjektiven Präferenzskalen ergeben:

1. AECB
2. BACD
3. BCAD
4. BCDA
5. CBDA
6. CDBA
7. DCBA

Empirisch besitzen wir indessen nur die sieben Präferenzskalen, nicht aber die Anordnung der Stimuli und Halbierungspunkte auf dem Kontinuum. Die Aufgabe ist, diese zu rekonstruieren. Zwei Präferenzskalen (1 und 7) zeigen unmittelbar die Anordnung der Stimuli auf dem Kontinuum. Von ihnen ausgehend können wir die Rangordnung der Halbierungspunkte feststellen. Nebeneinander liegende Befragtengruppen werden sich immer nur durch die umgedrehte Anordnung zweier nebeneinanderliegender Stimuli unterscheiden, z.B.: (ABCD) und (BACD). Das heißt aber, die Grenze zwischen diesen zwei Befragtengruppen stellt den Halbierungspunkt dar. Auf diese Art und Weise kann man sämtliche Halbierungspunkte feststellen.

Praktisch hat man also zuerst mit der Präferenzskala einer Extremgruppe begonnen, dann diejenige Gruppe gesucht, die

sich von ihr nur durch die Vertauschung zweier nebeneinander liegender Stimuli unterscheidet, daraufhin die nächste und so fort. Am Schluß dieser Prozedur ergibt sich daraus sowohl die Rangordnung der Stimuli, wie die der Halbierungspunkte und damit die der Befragten.

Bis hierher haben wir allerdings noch nicht mehr, als mit den kategorischen Methoden erreicht. Mit der komparativen Methode lassen sich jedoch noch weitere Informationen aus dem empirischen Material holen. Man kann die relative Größe der Abstände zwischen einigen der Stimuli bestimmen und somit schon zu einer quasi-metrischen Skalierung gelangen.

Gehen wir dazu wieder von Schnaubild 9 aus. Vier Stimuli können, je nach den relativen Abständen voneinander, verschiedene Präferenzskalen hervorrufen. Für vier Stimuli ergibt sich z.B. aus der Lage der Halbierungspunkte der zwei äußeren und der zwei mittleren Stimuli eine ganz bestimmte Information. Kommt H_{AD} vor H_{BC} , so folgt daraus, daß die Entfernung zwischen AB größer ist, als die zwischen CD. Für H_{BC} vor H_{AB} gilt das Gegenteil. Im ersten Fall hatten sich die Präferenzskalen (I) im zweiten Fall die (II) ergeben.

(I)	(II)
1. ABCD	1. ABCD
2. BACD	2. BACD
3. BCAD	3. BCAD
4. BCDA	4. CBAD
5. CBDA	5. CBDA
6. CDBA	6. CDBA
7. DCBA	7. DCBA

Die Rangordnung der zwei Halbierungspunkte gibt uns damit die zusätzliche (metrische) Information, daß

$$|S_B - S_A| > |S_D - S_C|$$

Dazu tritt die Information, die sich sowieso aus der Rangordnung der Stimuli ergibt, also etwa bei ABC:

$$|S_C - S_A| > |S_C - S_B|$$

$$|S_C - S_A| > |S_B - S_A|$$

Da nun jede Untergruppe von vier Stimuli Information über die relative Größe von zwei Entfernungen enthält, wird man im allgemeinen soviel metrische Information der geschilderten Art erhalten, wie Kombinationen von vier Stimuli möglich sind.

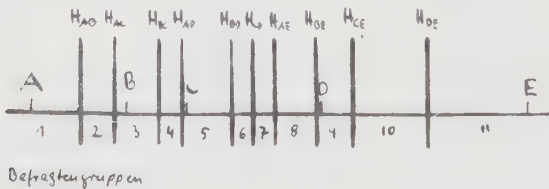
Ein Beispiel:¹⁾

Für fünf Stimuli ABCDE ergäben sich folgende Präferenzskalen:

1. ABCDE
2. BACDE
3. BCADE
4. CBADE
5. CDBAE
6. CDBAE
7. DCBAE
8. DCBEA
9. DCEBA
10. DECBA
11. EDCBA

Daraus ergibt sich die Rangordnung der Stimuli, Halbierungspunkte und Befragtengruppen:

Schaubild 11



1) Torgerson, W.S.: a.a.O., 1950, S.409.

Die erzielte metrische Information ist dabei:

Untergruppe	Rangordnung der Halbierungspunkte	metr. Inform.
ABCD	$H_{BC}^P H_{AD}$	$ s_B - s_A < s_D - s_C $
ABCE	$H_{BC}^P H_{AE}$	$ s_B - s_A < s_E - s_C $
ABDE	$H_{BD}^P H_{AE}$	$ s_B - s_A < s_E - s_D $
ACDE	$H_{CD}^P H_{AE}$	$ s_C - s_A < s_E - s_D $
BCDE	$H_{CD}^P H_{BE}$	$ s_C - s_D < s_E - s_D $

3. Partielle Präferenzskalen

a. Dieses Modell läßt sich vereinfachen. Die Befragten werden nicht beauftragt, alle n Stimuli nach ihrer Präferenz zu ordnen, sondern diejenigen r Stimuli aus der Gesamtzahl herauszusuchen, die ihrer Position auf dem Kontinuum am nächsten sind und nur diese auf einer Präferenzskala zu ordnen.¹⁾ Die Präferenzskalen sind jetzt also nicht mehr total sondern partiell.

An einem Beispiel: nehmen wir an $r=3$ und $n=5$. Das heißt, jeder Befragte soll die drei Items oder Stimuli aus fünf auswählen, mit denen er am ehesten übereinstimmt und ordnen. Würde er sie, ähnlich wie bei der Scalogramm-Analyse von Punktitems nur auswählen, so ergäbe sich folgende Anordnung der Items und Befragten.

1) siehe dazu Torgerson, W.S.: a.a.O., 1958, S.410 ff. und Coombs, C.H.: a.a.O., 1958.

Schaubild 12

Anforderung typen	Stimuli					Σ
	A	B	C	D	E	
I	x	x	x			N_I
II		x	x	x		N_{II}
III			x	x	x	N_{III}

Die Befragten lassen sich in drei Gruppen unterteilen und ordnen.

b. Weit mehr Information über die Verteilung der untersuchten Eigenschaft, geben uns die Präferenzskalen für die r Stimuli.

Schaubild 13

Anforderung typen	Stimuli					Σ
	A	B	C	D	E	
I	1	2	3			
II	2	1	3			
III	3	1	2			
IV		2	1			
V		1	1	3		
VI		3	2			
VII			2	1	3	
VIII			3	1	2	
IX			3			

Hier können die Befragten in elf geordnete Gruppen unterteilt werden. Spalten und Zeilen des Skalogramm müssen dabei so umarrangiert werden, daß

a. die Eintragungen in den Spalten und Zeilen ohne Leerstellen sind,

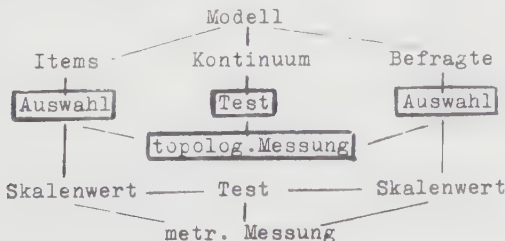
- b. die Eintragungen in der ersten Zeile von links nach rechts, und die in der letzten Zeile von rechts nach links monoton ansteigen,
- c. die Eintragungen in der ersten Spalte von oben nach unten und die in der letzten von unten nach oben zunehmen und
- d. die Eintragungen aller anderen Spalten zuerst ab- und dann zunehmen.

Sind diese Kriterien erfüllt, so ist erstens die Hypothese, des zugrundeliegenden eindimensionalen Kontinuum getestet (siehe Abschnitt I) und zweitens ist, wie aus der Gegenüberstellung der Schaubilder 12 und 13 ersichtlich, die erzielte Information umfangreicher, als im Fall einer einfachen Skalogramm-Analyse. Außerdem erhalten wir natürlich auch hier die weiter oben besprochene metrische Information. Voraussetzung ist dafür, daß $r \geq 3$.

C. Zusammenfassung

1. Ergebnisse der topologischen Messung

a. In Schaubild 4 hatten wir eine schematische Darstellung unseres Modells gegeben:



Die eingerahmten Elemente haben wir besprochen. Damit ist der erste Abschnitt unseres Modells abgeschlossen, Methoden und Voraussetzungen für eine topologische Messung sind bekannt.

Wir haben dabei gezeigt, wie sich aus der Zahl der möglichen Items die geeignetsten herausfinden lassen, wie diese einer repräsentativen Gesamttheit vorgelegt werden und wie wir dadurch zu einer topologischen Messung der zugrundeliegenden qualitativen Eigenschaft kommen können. Gleichzeitig mit dieser Messung haben wir eine Art von Test auf die Hypothese der Ein-dimensionalität aufgezeigt.

b. Was haben wir insgesamt damit erreicht? Wo liegt die Verbesserung gegenüber der traditionellen "direkten" Frage? Die Verbesserung ist dreifach:

Erstens können wir mit der hier beschriebenen Methode, unbeschadet größter Verzerrungen in den Fragen, Items oder Stimuli eine scharfe Trennung von "Pro" und "Kontra" treffen, damit einen invarianten "G-Punkt" finden.

Mit anderen Worten, wir können relativ exakt die Befragten und damit die Gesamttheit in zwei Gruppen trennen: diejenigen, die eine prinzipiell positive Einstellung zu dem betreffenden Untersuchungsobjekt haben und diejenigen, deren Eigenschaft prinzipiell negativ ist. Das war bisher in dieser eindeutigen Art und Weise nicht möglich.

Zweitens, zu dieser Trennung in zwei Gruppen erhalten wir eine ungefähre Vorstellung von der Verteilung der Einstellung auf die Befragten. Diese Verteilung ist allerdings nur auf einer Rangskala gegeben. Wir können jedoch daraus entnehmen,

ob sich die Träger der positiven und negativen Einstellung jeweils am Ende der Skala konzentrieren, also die Meinungen extrem vertreten werden, oder ob die Verteilung der Positionen der Befragten um die Mitte des Kontinuum am dichtesten sind, sich also die "Pro" und "Kontra" nicht allzu scharf voneinander unterscheiden.

Drittens, wir sind in der Lage, einen Test auf die Ein-dimensionalität des zugrundeliegenden Kontinuum durchzuführen. Das heißt, wir können feststellen, ob für eine bestimmte Einstellung eine oder mehrere Variable verantwortlich sind. Wenn wir, um ein etwas banales Beispiel zu bringen, die "Einstellung zur UNO" testen wollen und herausfinden, daß diese Einstellung nicht skalierbar ist, so wissen wir, daß die Befragten zur UNO eben nicht einfach positiv oder negativ stehen, sondern sie in gewissen Aspekten befürworten, in anderen ablehnen. Diese Aspekte oder Dimensionen kann man nun einzeln untersuchen.

Was dies für die Marktforschung analog angewendet bedeutet, braucht nicht ausgeführt zu werden. Wir sind dadurch in der Lage herauszufinden, welche Variable eines Produkts wie gestaltet werden müssen, um einen maximalen Verkaufserfolg zu erzielen. Testet man z.B. die Einstellung zu einer bestimmten Seifensorte, so wird sich zuerst wahrscheinlich keine Skalierbarkeit ergeben. Der Grund ist das Vorhandensein mehrerer Dimensionen. Die Seifenproben (die hier unser Stimuli darstellen) können im Geruch, in der Farbe, in der Schaumfähigkeit usw. variieren. Testet man jetzt jede dieser Dimensionen einzeln und erhält die Verteilung der Einstellung auf die einzelnen Dimensionen, so wird man angeben können, welche Kombination

der betreffenden Variablen den besten Verkaufserfolg hervorgerufen wird.

Die Bedeutung dieser drei Resultate ist evident. Sie bringen eine wesentliche Verbesserung der bestehenden "direkten" oder quasi-intuitiven Befragungsmethoden. Das hier vorgeführte Modell bedeutet damit die Einführung relativ exakter Methoden in das von der Statistik bisher etwas vernachlässigte Gebiet der Markt- und Meinungsforschung.

2. Weitere Aufgaben

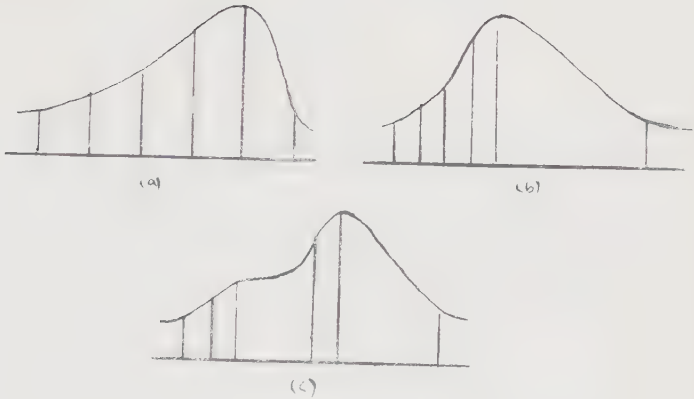
a. Die bisher behandelten Elemente unseres Modells erlauben jedoch nur eine topologische Messung der untersuchten Eigenschaften. Das heißt, wir können nur die Verteilung der Eigenschaft auf einer Rangskala feststellen. Wenn wir beispielsweise sechs Items oder Stimuli hätten, so könnten wir folgendes Ergebnis erhalten: A: 8%, B: 12%, C: 15%, D: 25%, E: 30%, F: 10%. Das bedeutet, die Einstellung von 3% der Befragten wird durch Stimulus A gekennzeichnet usw.

Schaubild 14



Damit wissen wir aber noch wenig über die wahre zugrundeliegende Verteilung. Da die Items nur eine Rangskala darstellen, kann sie so, wie in Schaubild 15(a), (b), (c) oder anders aussehen.

Schaubild 15



Die zuletzt besprochenen komparativen Methoden liefern uns zwar schon darüber hinausgehende Informationen, die Ergebnisse sind teil-metrisch, eine echte metrische Messung liegt aber auch hier nicht vor.

b. Wenn wir jedoch wirklich exakt messen wollen, speziell wenn es auf den Vergleich der Einstellung zu verschiedenen Komponenten eines Objekts ankommt (wie oben im Ergebnis 3 angedeutet), so wird eine genaue Kenntnis der Verteilung und damit eine metrische Messung immer notwendiger. Das heißt jedoch, wir müssen den Skalenwert unserer Items oder Stimuli kennen. Unsere Information darf also nicht nur

A (1) B (1) C (1) D (1) E

sondern auch

$$A - B = h$$

$$B - C = i$$

$$C - D = j$$

$$\begin{matrix} \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{matrix}$$

umschließen, wobei $h, i, j \dots$ bestimmte Werte annimmt. Der Abstand zwischen den Positionen der einzelnen Items oder Stimuli auf dem Kontinuum muß also bekannt sein, erst dann kann dieser, jetzt metrische, Maßstab auch zur metrischen Messung der Einstellung selbst dienen.

Schaubild 16



Die Aufgabe des nächsten Abschnitts unseres Modells wird es mithin sein, Methoden für eine derartige metrische Messung zu finden.

Dies scheint ein etwas absurdes Unterfangen, bedeutet es doch praktisch die Feststellung: dieses Item ist doppelt so positiv zum Untersuchungsobjekt wie jenes, diese Seifensorte duftet doppelt so stark wie jene, das Grün dieser Packung ist doppelt so intensiv wie das jener usw. Trotzdem können unter der Voraussetzung, daß die untersuchte Eigenschaft überhaupt einen Skalenwert auf einer metrischen Skala besitzt, Aussagen dieser Art möglich werden.

Ist dies aber einmal geschehen, so können wir den gewonnenen metrischen Maßstab an die einzelnen Befragten anlegen und auch deren Position auf der metrischen Skala bestimmen. Das würde dann aber heißen, daß wir die Verteilung der untersuchten Eigenschaft auf die Befragten eindeutig bestimmen können. Doch dazu im nächsten Kapitel.

3. Kapitel: Die Konstruktion einer metrischen Skala für Meinun- =====gen und Einstellungen =====

Die einzige invariante Größe einer topologischen Skala war die Rangordnung der einzelnen Mengen der untersuchten Eigenschaft. Um eine metrische Abbildung der Mengen einer Eigenschaft auf die Menge der reellen Zahlen zu erreichen, muß ein weiteres Strukturmerkmal dieser Eigenschaft gegeben sein, welches wir über die Rangordnung hinaus, ebenfalls auf die Menge der reellen Zahlen isomorph abbilden können. Als dieses "zusätzliche Strukturmerkmal" hatten wir den "Abstand" zwischen zwei Items oder Stimuli erkannt.

Mit einer metrischen Skala wollen wir mithin im folgenden ganz allgemein eine Skala bezeichnen, für die sowohl die Rangordnung der einzelnen Mengen der betreffenden Eigenschaft, wie deren Abstände voneinander bestimmt sind. Eine solche metrische (oder Intervall-) Skala ist eindeutig bis auf lineare Transformationen oder Streckungen.

Wir werden in diesem Kapitel in Abschnitt A das Modell darlegen, daß der Bestimmung des Skalenwerts der einzelnen Items oder Stimuli zugrundeliegt. In Abschnitt B zeigen wir, wie man auf Grund dieses Modells praktisch zu der Bestimmung der Skalenwerte gelangen kann und bieten in Abschnitt C einen Test für den Grad der Anpassung des empirischen Materials auf das theoretische Modell. In Abschnitt D schließlich legen wir den so gewonnenen Maßstab wieder an die repräsentative Teil-

gesamtheit der Träger der zu untersuchenden Eigenschaft an und kommen so zur Bestimmung des Skalenwerts der Befragten und damit der exakten Verteilung der Eigenschaft.

A. Modell einer metrischen Skalierung von Stimuli

1. Der Grundgedanke des Modells

a. Für die Bestimmung des Skalenwerts der Items oder Stimuli benötigen wir abermals eine Beurteilergruppe, wie wir sie weiter oben schon bei der Auswahl der Items kennengelernt haben. Wir brauchen uns mit ihrer Zusammensetzung nicht nochmals näher zu beschäftigen, wichtig ist nur, daß auch hier wieder die Aufgabe der Befragten nicht in der Bekanntgabe ihrer Einstellung, sondern in einer möglichst objektiven Beurteilung der Items oder Stimuli liegt.

Als fundamentale Größe für die Konstruktion einer metrischen Skala dient die Wahrscheinlichkeit, daß von zwei dargebotenen Stimuli oder Items i, j , Stimulus j von den Beurteilern als der ranghöhere (größere, positivere, sympathischere, ...) bezeichnet wird. Auf dieser Basis läßt sich nach einigen vereinfachenden Annahmen eine metrische Skala konstruieren. Im Grunde stützen wir uns dabei auf nichts anderes als die bereits Fechner bekannte "Methode der richtigen und der falschen Fälle" oder Thurstones "Methode des paarweisen Vergleichs".¹⁾

b. Wir setzen voraus, jeder Stimulus besitze eine bestimmte Position auf einem eindimensionalen Kontinuum. Legt man nun diesen Stimulus einem Beurteiler vor, so wird er bei diesem

1) Pfanzagl, J.: a.a.O., 1959, S.33.

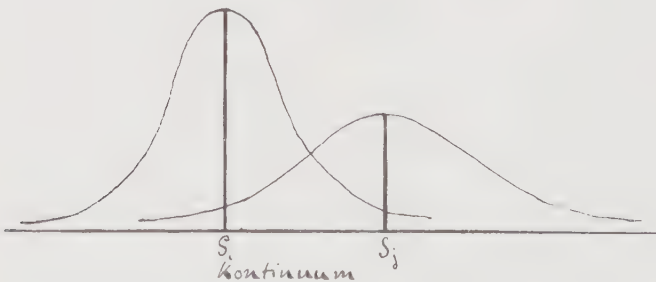
Torgerson, W.S.: a.a.O., 1958, S.155.

Fechner, G.T.: Elemente der Psychophysik, Leipzig, 1860, S.14.

Thurstone, L.L.: A law of comparative judgment, Psych.Rev., Bd.34, 1927, S.273-286.

einen Reaktionsprozeß hervorrufen, der sich auf den wahren Wert des Stimulus auf dem Kontinuum bezieht. Legt man den gleichen Stimulus einem Beurteiler mehrmals oder mehreren Beurteilern je einmal vor, so wird allerdings, infolge momentaner Fluktuationen in den Entscheidungszentren nicht immer die gleiche Reaktion erfolgen. Die einzelnen Reaktionen werden vielmehr so ausfallen, als ob sie sich auf Skalenwerte bezögen, die sich um den wahren Skalenwert verteilen. Wir nehmen nun an, daß diese Häufigkeitsverteilung normal sei. Damit aber gehört nun zu jedem wahren Skalenwert eines Stimulus eine normal verteilte Anzahl von Reaktionsprozessen. Der Mittelwert dieser Häufigkeitsverteilung ist ein Schätzwert für diesen wahren Skalenwert.

Schaubild 17



c. S_i und S_j seien die Skalenwerte zweier Stimuli, $\hat{\sigma}_i$ und $\hat{\sigma}_j$ die Streuung der entsprechenden Reaktionsprozesse. Diese zwei Stimuli legen wir nun mehreren Beurteilern vor, die angeben sollen, welches der beiden Stimuli das ranghöhere ist. Wäre das Modell streng deterministisch, so müßten sie immer, wenn

$S_j - S_i$ positiv ist Stimulus j als ranghöher bezeichnen und umgekehrt. Daraus könnten wir noch keine metrische Information schöpfen.

In Wirklichkeit werden die Vergleiche nicht von $S_j - S_i$ ausgehen, sondern von $R_j - R_i$, den Reaktionsprozessen, die für den N -ten Beurteiler mit den Stimuli j und i verbunden sind. Diese Differenz $R_j - R_i$ kann trotz $S_j > S_i$ negativ werden und zwar dann, wenn sich die Verteilungen von R_j und R_i überschneiden. Da sich R_j und R_i normal um S_j und S_i verteilen, wird sich $R_j - R_i$ ebenfalls normal verteilen und der Mittelwert dieser Verteilung einen Schätzwert für $S_j - S_i$, gemessen in $\sigma_{R_j - R_i}$ Einheiten, darstellen. Damit haben wir jedoch die gesuchte metrische Information gewonnen.

Schaubild 18

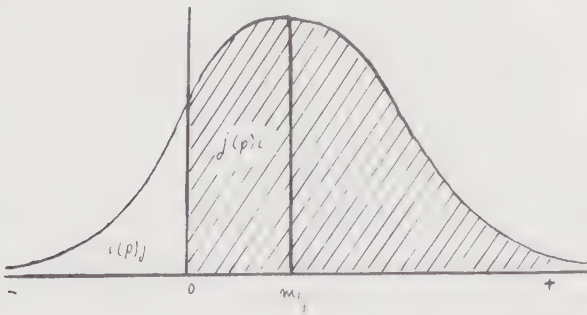


Schaubild 18 stellt die Verteilung aller $R_j - R_i$ dar. Das Verhältnis der schraffierten Fläche unter der Kurve der Normalverteilung zu der nichtschraffierten entspricht dem Verhältnis der Angaben " $j(p)_i$ " zu " $i(p)_j$ ". Der Mittel-

wert der Verteilung ist der Schätzwert für $S_j - S_i$. Das heißt

$$(15) \quad S_j - S_i = m_{ij} \sigma_{R_j - R_i}$$

da

$$(16) \quad \sigma_{R_j - R_i} = \sqrt{\sigma_i^2 + \sigma_j^2 - 2 r_{ij} \sigma_i \sigma_j}$$

ergibt sich

$$(17) \quad S_j - S_i = m_{ij} \sqrt{\sigma_i^2 + \sigma_j^2 - 2 r_{ij} \sigma_i \sigma_j}$$

wobei r_{ij} die Korrelation zwischen jeweiligen Werten von R_i und R_j ist.

Die Gleichung (17) stellt theoretisch den Grundgedanken der Methode für die Konstruktion einer metrischen Skala dar.

Mit anderen Worten, wir haben den Beurteilern eine relativ einfache Aufgabe gestellt, sie sollen nur angeben, ob Stimulus oder Item j größer, schöner, positiver etc ist als Stimulus i . Das Verhältnis der Antworten haben wir dann über die Verteilung der Reaktionsprozesse in Beziehung gesetzt zu der Differenz der Skalenwerte beider Stimuli. Diese Differenz ist aber das Strukturmerkmal einer metrischen Skala.

Gleichung (17) gibt uns nun $S_j - S_i$ als Funktion von $m_{i,j}$, σ_i , σ_j und r_{ij} . Empirisch ist uns allerdings nur das Verhältnis p_{ij} gegeben.

$$p_{ij} = \frac{\sum j(p)_{ij}}{\sum i(p)_{ij}}$$

Dabei ist

$j(p)_{ij}$ die Summe der Male, in denen Stimulus j als ranghöher bezeichnet wurde als Stimulus i

$i(p)j$ die Summe der Male, in denen Stimulus i als ranghöher bezeichnet wurde als Stimulus j

und

$j(p)i$ wenn $R_j - R_i < 0$

$i(p)j$ wenn $R_j - R_i > 0$

Von p_{ij} können wir zu m_{ij} gelangen. Damit ist dieses gegeben. Bei n Stimuli, n Skalenwerte, n Streuungen der Reaktionsprozesse und $n(n-1)/2$ r -Werten erhalten wir jedoch selbst dann, wenn wir Skalenwert und Streuung eines Stimulus willkürlich als 0-Punkt und Einheit definieren immer noch $2(n-1) + n(n-1)/2$ Unbekannte bei $n(n-1)/2$ Gleichungen. Wir werden daher vereinfachende Annahme einführen müssen, um das Gleichungssystem lösbar zu machen.

2. Vereinfachende Annahmen

Um die Zahl der Unbekannten so zu beschränken, daß unsere Beobachtungsgleichungen lösbar werden, könnten wir theoretisch eine große Zahl von vertretbaren Vereinfachungen einführen. Wir wollen hier nur die zwei einfachsten erwähnen.¹⁾

a. Die großzügigste Vereinfachung bestände darin, die Standardabweichung der Verteilung der Reaktionsprozesse für alle Stimulipaare als konstant anzunehmen

1) Siehe dazu auch Thurstones Fall III, IV, V des "law of comparative judgement", in denen ähnliche Vereinfachungen behandelt werden. Thurstone, L.L.: a.a.O., 1927; auch Torgerson, W.S.: a.a.O., 1958.

$$(19) \quad \sigma_j = \sigma_i = \sigma$$

Ferner sei ebenfalls $r_{i,j}$ konstant

$$(20) \quad r_{ij} = r$$

Daraus ergibt sich

$$(21) \quad S_j - S_i = m_{ij} \sqrt{2\sigma^2(1-r)}$$

das heißt, wenn wir setzen

$$(22) \quad c = \sqrt{2\sigma^2(1-r)}$$

$$(23) \quad S_j - S_i = c \cdot m_{ij}$$

Die Einheit der Skala kann nun so gewählt werden, daß $c = 1$.

Dann wird

$$(24) \quad S_j - S_i = m_{ij}$$

Wir werden später sehen, daß dieses Gleichungssystem mit Hilfe der Methode der kleinsten Quadrate gelöst werden kann.

b. Nicht ganz so restriktiv wäre eine zweite Möglichkeit der Vereinfachung. Es sei

$$(25) \quad \sigma_j - \sigma_i = d$$

wobei d sehr klein ist im Verhältnis zu σ_j und σ_i .

Dazu nehmen wir wie vorhin an, r_{ij} sei konstant

$$(20) \quad r_{ij} = r$$

In (17) eingesetzt ergibt sich dann

$$(28) \quad \delta_j - \delta_i = m_{ij} \sqrt{\sigma_j^2 + \sigma_i^2 + d^2 - 2\sigma_i \sigma_j (1-r)}$$

Da d sehr klein ist, können wir d^2 vernachlässigen und erhalten

$$(29) \quad \delta_j - \delta_i = m_{ij} \sqrt{\sigma_j^2 + \sigma_i^2 + 2\sigma_i \sigma_j - 2\sigma_i \sigma_j}$$

$$(30) \quad \delta_j - \delta_i = m_{ij} \sqrt{2\sigma_i \sigma_j (1+r)} \quad \sqrt{\sigma_i + d}$$

Da aber

$$(29) \quad \sqrt{\sigma_i + d} = \sqrt{\sigma_i} + \frac{d}{2\sqrt{\sigma_i}}$$

ergibt sich eingesetzt

$$(30) \quad \delta_j - \delta_i = m_{ij} \sqrt{2\sigma_i \sigma_j (1+r)} \left(\sqrt{\sigma_i} + \frac{d}{2\sqrt{\sigma_i}} \right)$$

Das heißt

$$(31) \quad \delta_j - \delta_i = m_{ij} \sigma_j \sqrt{2(1+r)} + \frac{1}{2} m_{ij} d \sqrt{2(1+r)}$$

$$(25) \quad \sigma_j - \sigma_i = d$$

$$(32) \quad \delta_j - \delta_i = m_{ij} \sqrt{2(1+r)} \left[\sigma_j + \frac{1}{2} (\sigma_j - \sigma_i) \right]$$

und endlich

$$(33) \quad \delta_j - \delta_i = m_{ij} (\sigma_i + \sigma_j) \cdot \sqrt{\frac{1}{2} (1+r)}$$

Auch dieses Gleichungssystem läßt sich, wie wir später sehen werden mit unseren empirischen Beobachtungswerten lösen.

3. Zusammenfassung.

Ausgehend von einem einfachen, topologischen Vergleich zweier Stimuli sind wir, über Annahmen bezüglich der Verteilung der die Ergebnisse dieses Vergleichs bestimmenden Reaktionsprozesse zu dem allgemeinen Gleichungssystem (18)

$$(18) \quad S_j - S_i = m_{ij} \sqrt{\sigma_j^2 + \sigma_i^2 - 2r_{ij} \sigma_i \sigma_j}$$

gekommen.

Dieses Gleichungssystem ist in der vorliegenden Form zur Bestimmung des Skalenwerts, damit zur Konstruktion einer metrischen Skala noch nicht geeignet. Die Zahl der Unbekannten überschreitet die Zahl der Gleichungen. Wir müssen daher vereinfachende Annahmen einführen. Die allgemeinste wäre, sowohl die Standardabweichungen wie die Korrelationskoeffizienten der Reaktionsprozesse als konstant anzunehmen. Damit kommen wir zu dem Gleichungssystem (24)

$$(24) \quad S_j - S_i = m_{ij}$$

Hier ist die Zahl der Gleichungen größer als die der Unbekannten, wir werden uns also zur Lösung des Gleichungssystems der Methode der kleinsten Quadrate bedienen müssen.

Eine zweite, etwas gemäßigtere Vereinfachung wäre folgende: $r_{i,j}$ wird zwar weiterhin als konstant angesehen, die Standardabweichungen der Reaktionsprozesse sind jetzt indessen verschieden, aber nur geringfügig voneinander abweichend. Unter diesen Annahmen gelangen wir zu dem Gleichungssystem (33)

$$(33) \quad S_j - S_i = m_{ij} (\sigma_j + \sigma_i) \sqrt{\frac{1}{2} (1-r)}$$

das ebenfalls, wie wir sehen werden, lösbar ist.

Im folgenden wollen wir nun das analytische Vorgehen bei der Konstruktion der metrischen Skala im einzelnen besprechen. Zuerst für den Fall des Gleichungssystems (24), dann für den

24. 1911

2. Das empirische Material

1. Das empirische Material

a. Das für die Konstruktion unserer Skala vorhandene empirische Material wird aus einer großen Zahl von paarweisen Vergleichen zweier Stimuli bestehen. Jeder Beurteiler vergleicht jeden Stimulus mit jedem anderen. Seine Aufgabe besteht dabei darin anzugeben, welcher Stimulus des jeweiligen Paares positiver, größer, schöner usw. ist - je nach der untersuchten Einstellung. Die Befragten werden angewiesen, immer einen der Stimuli als ranghöher zu bezeichnen. Beurteilungen von der Form: "Stimulus i ist gleich Stimulus j" oder "ich weiß nicht" sind nicht zugelassen.

Die Zahl der Beurteiler sollte möglichst groß sein. Je mehr Vergleiche für ein bestimmtes Paar vorhanden sind, umso exakter werden die Ergebnisse sein. Die Zusammensetzung der Beurteilergruppe braucht nicht einem "pilot sample" der später zu befragenden Beurteiler entsprechen. Es sollte sich vielmehr um eine Reihe

besonders urteilsfähiger Personen handeln, die möglichst keine extreme eigene Einstellung zu dem Untersuchungsobjekt besitzen. Die bereits bei der Auswahl der Items beschäftigten Beurteiler könnten hier eventuell wieder eingesetzt werden.

Die Stimulipaare selbst sollten den Beurteilern möglichst in zufälliger Reihenfolge präsentiert werden. Um Verzerrungen zu vermeiden, ist darauf zu achten, daß Paare, die einen gemeinsamen Stimulus besitzen, möglichst nicht direkt nacheinander vorgelegt werden. Es sollte auch für den Beurteiler keine Systematik in der Schwierigkeit der Beurteilung oder der Reihenfolge der "richtigen" Antworten zu entdecken sein.

Noch ein Wort zu dem Begriff des "Stimulus". Wollten wir streng im Rahmen unseres eingangs entworfenen Gesamtmodells der Skalierung von Meinungen und Einstellungen bleiben, so würden die Stimuli nur Items umfassen. Und zwar diejenigen Items, die wir nach den in Kapitel 2 Abschnitt A beschriebenen Methoden ausgewählt haben. Unsere n Items ergäben dann $n(n-1)/2$ Itempaare, die den Beurteilern vorgelegt werden. Wir könnten also in diesem Fall überall statt des Ausdrucks "Stimulus" das vertraute "Item" weiter benutzen. Die hier vorgeführte Methode der metrischen Skalierung läßt sich jedoch auch auf andere Stimuli als Items anwenden. So könnten wir, ähnlich wie bei dem oben vorgeführten komparativen Modell, Warenproben auf diese Art skalieren. Zum Beispiel eine Reihe von Speiseeisproben verschiedener Süße, verschieden stark duftende Seifen, verschiedene Verpackungen eines Produkts usw. Das ist durchaus möglich, vorausgesetzt nur, wir achten darauf, daß die vorgelegten Proben jeweils nur in einer Dimension variieren. Die

Einstellung zu gewissen Produkten etwa kann also auch über andere Stimuli als Items gemessen werden. Wir erweitern hier, wie schon oben, bei der Besprechung der komparativen Methode unser Modell und sprechen allgemein von der Skalierung von Stimuli nicht nur von Items. Da für andere Stimuli als Items - speziell beim Warentest - sowieso nur eine beschränkte Zahl von Proben zur Verfügung stehen werden, tritt das Problem der Auswahl hier nicht auf.

b. Wir besitzen nun eine Reihe von ausgewählten Items oder sonstigen Stimuli, etwa Warenproben. Diese werden von einer möglichst großen Beurteilergruppe paarweise miteinander verglichen. Diese geben dabei jeweils an, welches der beiden Stimuli das ranghöhere ist. Wie werden die so ermittelten Informationen ausgewertet?

Die aggregierten Beurteilungen der $n(n-1)/2$ Stimulipaare bilden in Form der Häufigkeitsmatrix (H-Matrix) unser Urmaterial.

H-Matrix

	Stimulus j ($j = 1, 2, 3, \dots$)			
	1	2	3	...
1	-	h_{12}	h_{13}	h_{14}
2	h_{21}	-	h_{23}	h_{24}
3	h_{31}	h_{32}	-	h_{34}
4	h_{41}	h_{42}	h_{43}	-

Das Element h_{ij} bezeichnet dabei die Häufigkeit, mit der Stimulus i als ranghöher bezeichnet wurde. Die Diagonalzellen der Matrix bleiben leer, die Stimuli werden nicht mit sich selbst verglichen.

Ausgehend von der H-Matrix können wir eine zweite Matrix, die P-Matrix konstruieren. p_{ij} bezeichnet jetzt den Anteil der Male, in denen Stimulus i als ranghöher bezeichnet wurde. Wenn N die Gesamtzahl der Beurteiler ist, so ergibt sich p_{ij} ganz einfach aus

$$(34) \quad p_{ij} = \frac{1}{N} h_{ij}$$

und

$$(35) \quad p_{ji} = 1 - p_{ij}$$

P-Matrix

	Stimulus j ist			
	1	2	...	n
1	-	p_{12}	p_{1j}	p_{1n}
2	p_{21}	-	p_{2j}	p_{2n}
...				
i	p_{i1}	p_{i2}	p_{ij}	p_{in}
...				
n	p_{n1}	p_{n2}	p_{nj}	-

Die Diagonalzellen bleiben wieder unbesetzt, oder werden mit 0,5 ausgefüllt. Das Ergebnis bleibt dadurch unverändert.

Die P-Matrix schließlich transformieren wir in eine M-Matrix, in der m_{ij} den p_{ij} , entsprechenden Mittelwert der Normalverteilung bezeichnet. Die Werte von m_{ij} können einer Tabelle entnommen werden. m_{ij} wird positiv für alle Werte von p_{ij} sein, die über 0,5 liegen und negativ für alle p_{ij} unter 0,5. p_{ij} -Werte von 1,0 oder 0,0 können nicht verwendet werden, weil die entsprechenden m_{ij} -Werte unverhältnismäßig hoch waren. Die entsprechenden Zellen bleiben leer. Ebenso werden die Diagonalzellen leer bleiben.

Diese m_{ij} -Werte sind nun Schätzungen der Differenz $S_j - S_i$ gemessen in $\hat{\sigma}_{R,R}$ Einheiten.

	m_{12}	m_{13}	m_{14}
m_{21}			
m_{31}			
m_{41}			

Jede dieser Schätzungen wird von der wahren Differenz $S_j - S_i$ abweichen. Wir werden daher Verfahren aufzeigen müssen, die es erlauben, die Summe dieser Abweichungen zu minimieren und uns damit relativ gute Schätzwerte für die Unbekannten unserer

Gleichungen liefern. Das soll die Aufgabe des nächsten Abschnitts sein.

2. Bestimmung des Skalenwerts (I)

Wir wollen zuerst die Berechnung des Skalenwerts unter den vereinfachenden Annahmen der Gleichung (24) behandeln.

$$(24) \quad m_{ij} = S_j - S_i$$

a. Praktisch ist die Bestimmung dieser Skalenwerte sehr einfach. Das arithmetische Mittel jeder Spalte oder Zeile der M -Matrix gibt uns den Skalenwert des jeweiligen Stimulus, ausgedrückt als Abweichung vom Mittel aller Skalenwerte.

$$(36) \quad S_j = \frac{1}{n} \sum_{i=1}^n m_{ij} \quad (j = 1, 2, \dots, n)$$

Da wir vorerst den 0-Punkt der Skala völlig willkürlich setzen, können wir zu allen erhaltenen Skalenwerten den des höchsten negativen addieren. Dadurch werden alle positiv und sind einfacher zu handhaben. Wir werden diesen Vorgang später an Hand eines Beispiels klarer machen.

Damit haben wir das gewünschte Ziel erreicht. Unsere Items oder Stimuli sind nicht nur rangmäßig geordnet, sondern auch der jeweilige Abstand zwischen zwei Stimuli ist bekannt. Das heißt, statt eines topologischen besitzen wir jetzt einen metrischen Maßstab für die zu untersuchende Einstellung.

Gleichung (36) steht allerdings bis jetzt noch als leere Behauptung da. Wir werden daher zuerst den Beweis für die

Berechtigung dieses Vorgehens liefern müssen.¹⁾ Dazu im Folgenden.

b. Zunächst ein Wort zu den Bezeichnungen: S_i und S_j sowie m_{ij} beziehen sich im folgenden nicht auf die theoretischen, wahren Skalenwerte, bzw. Differenzen zwischen den Skalenwerten, sondern auf die Schätzwerte, die wir für sie aus dem empirischen Material gewonnen haben. Mit m'_{ij} hingegen wollen wir abgeleitete Werte für die Differenzen bezeichnen. Also: die m_{ij} werden benutzt, um die S_i und S_j zu bestimmen, aus den errechneten Schätzwerten kann man nun wieder ein m'_{ij} ableiten.

$$(37) \quad m'_{ij} = S_j - S_i$$

Hätten wir fehlerfreies Material, so müßte $m_{ij} = m'_{ij}$ sein. In Wirklichkeit werden sich jedoch Abweichungen ergeben.

Unsere Aufgabe ist es daher, diejenigen Skalenwerte zu bestimmen, für die das Quadrat der Abweichungen $m_{ij} - m'_{ij}$ ein Minimum wird.

$$(38) \quad A = \sum_{i=1}^n \sum_{j=1}^n (m_{ij} - m'_{ij})^2 \rightarrow \text{Minimum}$$

Da

$$m'_{ij} = S_j - S_i$$

ergibt sich

$$(39) \quad A = \sum_{i=1}^n \sum_{j=1}^n (m_{ij} - S_j + S_i)^2$$

1) Vgl.: Mosteller, F.: a.a.O., 1951, S.3-16.

Um A zu minimieren müssen wir nach jedem S partiell ableiten.

Matrix der quadratischen Abweichung

	stimulus j								
	1	2		j					n
1	0	$(m_{12} - S_2 + S_1)^2$		$(m_{1j} - S_j + S_1)^2$				$(m_{1n} - S_n + S_1)^2$	
2	$(m_{21} - S_1 + S_2)^2$	0		$(m_{2j} - S_j + S_2)^2$				$(m_{2n} - S_n + S_2)^2$	
...	...	...		...				...	
i	$(m_{i1} - S_1 + S_i)^2$	$(m_{i2} - S_2 + S_i)^2$		$(m_{ij} - S_j + S_i)^2$				$(m_{in} - S_n + S_i)^2$	
...	...	...		...				...	
n	$(m_{n1} - S_1 + S_n)^2$	$(m_{n2} - S_2 + S_n)^2$		$(m_{nj} - S_j + S_n)^2$				0	

Da jedes S jeweils nur in einer Zeile und Spalte der Matrix erscheint und ferner

$$(40) \quad m_{ij} = -m_{ji} \quad \text{und} \quad (S_j - S_i) = -(S_i - S_j)$$

brauchen wir uns nur mit den Spalten zu beschäftigen. Die Elemente einer Spalte werden in Bezug auf S_j differenziert, es ergibt sich

$$(41) \quad \frac{dA}{dS_j} = -2 \sum_{i=1}^n (m_{ij} - S_j + S_i)^2 \quad (j = 1, 2, \dots, n)$$

Für $\frac{dA}{dS_j}$ gleich Null

$$(42) \quad \sum_{i=1}^n S_j = \sum_{i=1}^n m_{ij} + \sum_{i=1}^n S_i \quad (j = 1, 2, \dots, n)$$

Dividieren wir durch n

$$(43) \quad S_j = \frac{1}{n} \sum_{i=1}^n m_{ij} + \frac{1}{n} \sum_{i=1}^n S_i \quad (j = 1, 2, \dots, n)$$

und bestimmen als Ursprung das Mittel aller Skalenwerte

$$(44) \quad \frac{1}{n} \sum_{i=1}^n S_i = 0$$

so ergibt sich

$$(45) \quad S_j = \frac{1}{n} \sum_{i=1}^n m_{ij}$$

womit bewiesen ware, daß das einfache arithmetische Mittel der Spalten der M-Matrix Skalenwerte liefert, für die die quadratische Abweichung der ursprünglichen von den abgeleiteten Differenzen ein Minimum wird.

c. Weiter oben hatten wir gesagt, daß diejenigen Zellen der M-Matrix, die p_{ij} Werten von 0,0 oder 1,0 entsprechen, leer bleiben müssen. Diese Leerzellen werden wir immer dann erhalten, wenn sich die Verteilungen der Reaktionsprozesse um S_i und S_j nicht überschneiden. Wenn also die betreffenden Items oder Stimuli auf dem Kontinuum weit voneinander entfernt liegen.

Eine M-Matrix, die derartige Leerzellen enthält, können wir nicht in der oben beschriebenen Weise zur Bestimmung des Skalenwerts verwenden. Wir müssen daher ein anderes Verfahren benutzen.

Setzen wir zuerst den Fall, wir hätten eine vollständige M-Matrix und würden die Werte einer Spalte jeweils von den Werten der danebenliegenden Spalte abziehen. Es ergäbe sich dann:

$$\begin{aligned}
 m_{21} - m_{11} &= (S_2 - S_1) - (S_1 - S_1) = S_2 - S_1 \\
 m_{12} - m_{12} &= (S_2 - S_2) - (S_1 - S_2) = S_2 - S_1 \\
 m_{23} - m_{13} &= (S_2 - S_3) - (S_1 - S_3) = S_2 - S_1 \\
 m_{24} - m_{14} &= (S_2 - S_4) - (S_1 - S_4) = S_2 - S_1
 \end{aligned}
 \tag{46}$$

$$m_{2n} - m_{1n} = (S_2 - S_n) - (S_1 - S_n) = S_2 - S_1$$

oder

$$m_{2i} - m_{1i} = (S_2 - S_i) - (S_1 - S_i) = S_2 - S_1$$

Allgemein könnten wir schreiben

$$S_i - S_{i-1} = \frac{1}{n} \sum_{j=1}^n (m_{ij} - m_{i-1,j}) \quad (i = 1, 2, \dots, n)$$

Wenn wir nun $S_1 = 0$ setzen, so ergibt sich

$$\begin{aligned}
 S_1 &= 0 \\
 S_2 &= \frac{1}{n} \sum_{j=1}^n (m_{2j} - m_{1j}) \\
 S_3 &= S_2 + \frac{1}{n} \sum_{j=1}^n (m_{3j} - m_{2j}) \\
 &\vdots \\
 S_n &= S_{n-1} + \frac{1}{n} \sum_{j=1}^n (m_{nj} - m_{n-1,j})
 \end{aligned}
 \tag{47}$$

Auf diese Art und Weise kann Schritt für Schritt ein Skalenwert nach dem andern bestimmt werden. Wenn jetzt die Matrix unvollständig ist, so können wir einige der Gleichungen (46) nicht schreiben. Das wird dann der Fall sein, wenn in der gleichen Zeile zweier nebeneinanderliegender Spalten der M-Matrix keine m_{ij} Werte vorhanden sind. Ohne allzu großen Verlust an Genauigkeit können wir dann n nicht als die Zahl der Stimuli sondern als die Zahl der Gleichungen, damit der vorhandenen Differenzen betrachten. Auf diese Art und Weise erhalten wir auch für eine unvollständige Matrix relativ exakte Skalenwerte. Auch für dieses Vorgehen wollen wir später ein praktisches Beispiel bringen.

Würden wir dieses Verfahren auf eine vollständige Matrix anwenden, müßten sich natürlich die gleichen Skalenwerte ergeben, wie bei Anwendung der weiter oben beschriebenen Methode der Durchschnittsbildung.

d. Wir können jetzt zusammenfassen: Wenn wir das eingangs geschilderte theoretische Modell der Konstruktion einer metrischen Skala so vereinfachen, daß sich die Beziehung

$$(24) \quad S_j - S_1 = m_{1j}$$

ergibt, dann können wir den Skalenwert der einzelnen Items oder Stimuli auf zwei verschiedene Arten bestimmen:

1. Wir bilden den Durchschnitt jeder Spalte der M-Matrix

$$(36) \quad \bar{S}_j = \frac{1}{n} \sum_{i=1}^n m_{ij} \quad (j = 1, 2, \dots, n)$$

Dieser Durchschnitt stellt denjenigen Skalenwert dar, der die quadratischen Abweichungen der m_{ij} von den m_{ij} Werten minimiert.

Dieses Verfahren läßt sich indessen nur auf vollständige Matrizen anwenden. Überschneiden sich die Verteilungen der Reaktionsprozesse nicht, so werden in der M-Matrix Leerstellen entstehen, sie wird also unvollständig sein. Auf derartige unvollständige Matrizen können wir das zweite Verfahren der Bestimmung des Skalenwerts anwenden. Hier werden die einzelnen Skalenwerte nicht als Abweichungen vom Mittel aller Skalenwerte, sondern kumulativ auf Grund der Differenzen zwischen zwei nebeneinanderliegenden Stimuli berechnet.

Ehe wir nun Beispiele zu diesen Verfahren bringen können, müssen wir uns mit der Bestimmung des Skalenwerts unter den weniger restriktiven Annahmen der Gleichung (33) befassen.

3. Bestimmung des Skalenwertes (II)

a. Nicht immer wird die sehr generelle Annahme $G_j = G_i = G$ zu halten sein. Besonders wenn die Stimuli relativ komplex sind, werden die Verteilungen der Reaktionsprozesse nicht nur von der untersuchten Variablen abhängen. Wir werden daher hier annehmen, daß sich die einzelnen Verteilungen wenigstens etwas voneinander unterscheiden.

In diesem Fall müssen wir mit Gleichung (33) arbeiten.

$$(33) \quad S_j - S_i = m_{ij} (G_i + G_j) \sqrt{\frac{1}{2} (1+r)}$$

Wobei, so hatten wir gesagt, r konstant ist und die Differenz

zwischen $\hat{\sigma}_i$ und $\hat{\sigma}_j$ nur gering.¹⁾

Setzen wir nun

$$(4) \quad c = \sqrt{\frac{2}{1-r}}$$

so ist

$$(5) \quad c(S_j - S_i) = m_{ij} (\hat{\sigma}_i + \hat{\sigma}_j)$$

Für $i, j+1$ als Beobachtung

$$(6) \quad c(S_{j+1} - S_i) = m_{ij+1} (\hat{\sigma}_{j+1} + \hat{\sigma}_i)$$

Subtrahiert man (5) von (6) so erhält man

$$c(S_{j+1} - S_j) = m_{ij+1}(\hat{\sigma}_{j+1} + \hat{\sigma}_i) - m_{ij}(\hat{\sigma}_j + \hat{\sigma}_i)$$

oder

$$(53) \quad m_{ij+1} = m_{ij} \frac{\hat{\sigma}_j + \hat{\sigma}_i}{\hat{\sigma}_{j+1} + \hat{\sigma}_i} + \frac{c(S_{j+1} - S_j)}{\hat{\sigma}_{j+1} + \hat{\sigma}_i}$$

Die Beziehung zwischen m_{ij} und $m_{i,j+1}$ ist linear bis auf Veränderungen in $\hat{\sigma}_i$. Setzen wir $\hat{\sigma}_i = 1$, so können wir Gleichung

(53) schreiben:

$$(54) \quad m_{ij+1} = m_{ij} \frac{\hat{\sigma}_j + 1}{\hat{\sigma}_{j+1} + 1} + \frac{c(S_{j+1} - S_j)}{\hat{\sigma}_{j+1} + 1}$$

• 1) Thurstone bringt in seinem "case IV" die Lösung eines ähnlichen Gleichungssystems. Allerdings nimmt er eine Korrelation von 0 an. Wir werden ihm hier im wesentlichen folgen.
L.I. Thurstone: a.a.O., 1927, S.273-286. Im übrigen siehe
W.S. Torgerson: a.a.O., 1958, S.179-185.

Gleichung (54) stellt dann eine Gerade mit der Neigung

$$\frac{\sigma_j + 1}{\sigma_{j+1} + 1}$$

dar. Andererseits ist uns aber auch die Neigung der Geraden durch das Verhältnis der mittleren quadratischen Abweichungen v_j und v_{j+1} von m_{ij} und $m_{i,j+1}$ gegeben

$$(55) \quad \begin{aligned} v_j &= \frac{1}{n} \sum_{i=1}^n (m_{i,j} - \bar{m}_{\cdot j})^2 \\ v_{j+1} &= \frac{1}{n} \sum_{i=1}^n (m_{i,j+1} - \bar{m}_{\cdot j+1})^2 \end{aligned}$$

Es ergibt sich also approximativ

$$(56) \quad \frac{v_{j+1}}{v_j} = \frac{\sigma_j + 1}{\sigma_{j+1} + 1}$$

Allgemein können wir schreiben

$$(57) \quad v_j (\sigma_j + 1) = P$$

wobei P eine Proportionalitätskonstante ist.

Über j summiert ergibt sich nun

$$(58) \quad P \sum_{j=1}^n \frac{1}{v_j} = \sum_{j=1}^n \sigma_j + n$$

Jetzt können wir die Einheit der Skala so definieren, daß das durchschnittliche σ_j gleich eins ist.

$$(59) \quad \frac{1}{n} \sum_{j=1}^n \sigma_j = 1$$

So ergibt sich

$$(60) \quad P = \frac{2n}{\sum_{j=1}^n \frac{1}{v_j}}$$

Damit erhalten wir G_j , die Streuungen der Reaktionsprozesse um den Skalenwert des einzelnen Stimulus.

$$(61) \quad G_j = \frac{P}{v_j} - 1 \quad (j = 1, 2, \dots, n)$$

Nun können die Skalenwerte selbst berechnet werden. Als Ursprung der Skala definieren wir wieder das arithmetische Mittel aller Skalenwerte, so daß

$$(62) \quad \frac{1}{n} \sum_{i=1}^n S_i = 0$$

Da wir ohne Verlust metrischer Information die Skala beliebig strecken können, setzen wir

$$(63) \quad C = \sqrt{\frac{2}{1-r}} = 1$$

Somit ergibt sich für (33) bei Summierung über i

$$(64) \quad S_j = \frac{1}{n} \left(G_j \sum_{i=1}^n m_{ij} + \sum_{i=1}^n G_i m_{ij} \right) \quad (j = 1, 2, \dots, n)$$

Dabei ist

$$(65) \quad \sum_{i=1}^n G_i m_{ij} = - \sum_{j=1}^n G_j m_{ij}$$

Damit können wir die einzelnen S_j näherungsweise bestimmen.

Bei der praktischen Durchführung erfordert dieses Vorgehen allerdings erheblichen rechnerischen Aufwand (siehe das Beispiel S. 104). In dieser Hinsicht ist es unserem ersten Verfahren weit unterlegen. Außerdem kann es nicht angewendet werden, wenn die M -Matrix Leerzellen enthält, da die Schätzwerte für v_j alle auf den gleichen i Werten basieren müssen.

b. In diesem Fall müssen die einzelnen Skalenwerte wieder sukzessive angenähert werden. Wir wollen das Verfahren hier nur kurz andeuten:¹⁾

Wir ordnen zuerst die Spalten der M -Matrix ungefähr nach der Rangordnung der Stimuli. Dann bestimmen wir graphisch das Verhältnis

$$(66) \quad K_{i,j+1} = \frac{S_j + 1}{S_{j+1} + 1}$$

indem wir durch die Punkteschar $(m_{ij}, m_{i,j+1})$ jeweils eine Gerade legen und deren Neigung bestimmen.

Es ist dann

$$(67) \quad S_{j+1} = \frac{1}{K_{i,j+1}} (S_j + 1) - 1$$

Setzen wir nun willkürlich $S_1 + 1 \equiv a$, so ist

1) W.S. Torgerson, a.a.O., 1958, S.182.

$$\begin{aligned}
 \beta_1 &= a - 1 \\
 (68) \quad \beta_2 &= \frac{a}{k_{12}} - 1 \\
 \beta_3 &= \frac{a}{k_{12} k_{23}} - 1
 \end{aligned}$$

$$\beta_n = \frac{a}{k_{12} k_{23} k_{34} \dots k_{n-1, n}} - 1$$

oder allgemein

$$(69) \quad \beta_j = \frac{a}{k_{12} k_{23} k_{34} \dots k_{j-1, j}} - 1$$

Summiert und durch n dividiert ergibt sich

$$(70) \quad \frac{1}{n} \sum_{j=1}^n \beta_j = \frac{a}{n} \sum_{j=1}^n \frac{1}{k_{12} k_{23} k_{34} \dots k_{j-1, j}} - 1$$

Nun ist wieder

$$(71) \quad \frac{1}{n} \sum_{j=1}^n \beta_j = 1$$

das heißt

$$(72) \quad \frac{a}{n} \sum_{j=1}^n \frac{1}{k_{12} k_{23} \dots k_{j-1, j}} = 2$$

und

$$(73) \quad a = \frac{2n}{\sum_{j=1}^n \frac{1}{k_{12} k_{23} \dots k_{j-1, j}}} \quad (k = 1, 2, \dots, n)$$

Damit können wir a bestimmen. $\hat{\sigma}_j$ ergibt sich dann aus den Gleichungen (68).

Erinnern wir uns an Gleichung (52), so ist für $C=1$

$$(74) \quad S_{j+1} - S_j = m_{i,j+1} (\hat{\sigma}_{j+1} + \hat{\sigma}_i) - m_{ij} (\hat{\sigma}_j + \hat{\sigma}_i)$$

Für die Differenz $(S_{j+1} - S_j)$ existieren soviele Schätzungen, wie besetzte Zellenpaare in den Spalten j und $j+1$ vorhanden sind. Bezeichnen wir die Zahl dieser Paare mit n_b , so ist ihr Durchschnitt der gesuchte Schätzwert für die Differenz

$$(75) \quad S_{j+1} - S_j = \frac{1}{n_b} \left[\sum_{i=1}^{n_1} m_{i,j+1} \hat{\sigma}_{j+1} + \sum_{i=2}^{n_2} m_{i,j+1} \hat{\sigma}_i - \sum_{i=1}^{n_1} m_{ij} \hat{\sigma}_j - \sum_{i=2}^{n_2} m_{ij} \hat{\sigma}_i \right]$$

Wir können jetzt wieder den Ursprung willkürlich beim rangniedrigsten Stimulus festsetzen und dann sukzessive einen Skalenwert nach dem anderen bestimmen.

c. Zusammenfassung. Auch die Gleichung (33), die mit der weniger restriktiven Annahme arbeitet, die Streuungen der Reaktionsprozesse seien verschieden - wenn auch nur in geringem Ausmaß, läßt sich näherungsweise lösen. Sowohl, wie wir gesehen haben, für den Fall einer vollständigen, wie den einer unvollständigen M-Matrix. Die der praktischen Bestimmung der Skalenwerte zugrundeliegenden Gleichungen sind dabei (64) bzw. (75)

$$(64) \quad S_j = \frac{1}{n} \left(\hat{\sigma}_j \sum_{i=1}^n m_{i1} + \sum_{i=2}^n \hat{\sigma}_i m_{ij} \right) \quad (j=1, 2, \dots, n)$$

$$(75) \quad S_{j+1} - S_j = \frac{1}{n_b} \left[\sum_{i=1}^{n_b} m_{i,j+1} \delta_{i,j+1} + \sum_{i=1}^{n_b} m_{i,j} \delta_{i,j} - \sum_{i=1}^{n_b} m_{i,j} \delta_{i,j} - \sum_{i=1}^{n_b} m_{i,j} \delta_{i,j} \right]$$

Es muß nochmals betont werden, daß beide Lösungen nur approximativ sind. Der Grund liegt dafür in dem etwas dubiosen Schritt von Gleichung (53) zu Gleichung (54).

Im folgenden wollen wir die verschiedenen Methoden der Bestimmung des Skalenwerts an einem praktischen Beispiel durchführen. Wir werden dabei sehen, daß das auf den ersten Blick etwas kompliziert wirkende Verfahren praktisch leicht durchzuführen ist.

4. Praktisches Beispiel.

a. Aus finanziellen Gründen war es nicht möglich, eine eigene Befragung nach der hier vorgeführten Methode zu veranstalten. Da unser Beispiel jedoch nur illustrativen Charakter haben soll, genügt es, das Datenmaterial einer in einem anderen Zusammenhang veranstalteten amerikanischen Umfrage gekürzt und vereinfacht zu benutzen. Die dort erhaltenen Werte werden uns als Urmaterial dienen, das wir sowohl für den Fall (I) wie den (II) auswerten wollen.

Die Umfrage wurde 1953 von Hill veranstaltet.¹⁾ Das Ziel war, eine Reihe von Items zu skalieren, die als Maßstab für die "Einstellung zum Krieg in Korea" dienen können. Wir wollen hier der Einfachheit halber nur fünf Items herausgreifen (eigene Übersetzung):

1) Entnommen aus A.L. Edwards, a.a.O., 1957, S.31 f.

1. Wir sollten unseren Verbündeten in Korea mehr Geld geben, wenn sie es brauchen.
2. Wenn wir jetzt unsere Truppen aus Korea zurückziehen, wird dadurch alles nur noch schlimmer.
3. Der Krieg in Korea mag nicht gerade die beste Art und Weise sein, um den Kommunismus aufzuhalten, aber wir konnten nichts anderes tun.
4. Wir müssen den Krieg in Korea unbedingt gewinnen, was auch immer es kosten mag.
5. Wenn wir in Korea kämpfen, so schützen wir damit die USA.

Die fünf Items sind offensichtlich rangmäßig geordnet, jeder, der Item 4 bejaht, wird auch Item 1,2,3 bejahen usw. Der verschiedene Grad an positiver Einstellung zum Krieg in Korea, den sie beinhalten, ist uns indessen nicht bekannt. Unser Ziel ist es daher, jedem der fünf Items einen Skalenwert zuzuordnen, der diesen Grad wiedergibt. Also, die Items metrisch zu skalieren.

Wie weiter oben geschildert, werden die einzelnen Items zuerst paarweise einer Beurteilergruppe präsentiert. Jeder der Beurteiler soll für jedes Itempaar angeben, welches der beiden Items eine positivere Einstellung zum Krieg in Korea wiedergibt. Aus den Antwortauflagen ergibt sich die H-Matrix

H - Matrix

Item	1	2	3	4	5
1	-	51	54	62	68
2	43	-	49	59	60
3	40	45	-	49	63
4	32	35	45	-	51
5	26	34	31	43	-

Da die Items bereits rangmäßig geordnet waren, brauchen Spalten und Zeilen der H-Matrix nicht nach den einzelnen Gesamtnäufigkeiten umgestellt werden.

Aus der H-Matrix ergibt sich nun die P-Matrix, wobei

$$p_{ij} = \frac{h_{ij}}{N}$$

N ist die Zahl der Beurteiler.

P-Matrix

Item	1	2	3	4	5
1	-	0,543	0,574	0,660	0,723
2	0,457	-	0,521	0,628	0,638
3	0,426	0,479	-	0,521	0,670
4	0,340	0,372	0,479	-	0,543
5	0,277	0,362	0,330	0,457	-
	1,566	1,756	1,904	2,266	2,574

Ausgehend von der P-Matrix schließlich bestimmen wir an Hand einer Tabelle der Flächen unter der Gaußschen Glockenkurve die m_{ij} -Werte.

M-Matrix

Item	1	2	3	4	5
1	-	0,108	0,187	0,412	0,592
2	-0,108	-	0,053	0,327	0,353
3	-0,187	-0,053	-	0,053	0,440
4	-0,412	-0,327	-0,053	-	0,108
5	-0,592	-0,353	-0,440	-0,108	-

Die M-Matrix bildet nun unser Ausgangsmaterial, das wir nach den verschiedenen, oben geschilderten Methoden auswerten wollen.

b. Fall (I). Da wir über eine vollständige Matrix verfügen, können wir die Skalenwerte über Gleichung (24) bestimmen.

Item	1	2	3	4	5
$\sum_{i=1}^5 m_{i,j}$	-1,299	-0,625	-0,253	-0,684	+1,493
$\frac{1}{5} \sum_{i=1}^5 m_{i,j}$	-0,250	-0,125	-0,051	+0,137	+0,299
$\frac{1}{5} \sum_{i=1}^5 m_{i,j}$ + 0,250	$S_1 =$ 0	$S_2 =$ 0,125	$S_3 =$ 0,199	$S_4 =$ 0,387	$S_5 =$ 0,549

Es werden also ganz einfach die Durchschnitte der einzelnen Spalten gebildet. Da uns das Problem eines rationalen 0-Punkts bis hierhin noch nicht beschäftigt, können wir den Ursprung der Skala willkürlich auf S_1 legen. Die Skala wird so etwas übersichtlicher. Das Verfahren der Bestimmung des Skalenwerts im Fall (I) ist also denkbar einfach.

c. Nehmen wir nun an $\hat{b}_i$ und $\hat{b}_j$ seien verschieden, also Fall (II), so wird der Rechenaufwand größer. Wir müssen zuerst nach Gleichung (55) v_j bestimmen, davon ausgehend nach (60) P und damit schließlich nach (61) $\hat{b}_j$. Folgende Schritte werden dabei zurückzulegen sein:

M-Matrix (II)

$\bar{x} + cmis$	1	2	3	4	5
1	—	0,108	0,187	0,412	0,592
2	-0,108	—	0,053	0,297	0,592
3	-0,187	-0,053	—	0,053	0,440
4	-0,412	-0,297	-0,053	—	0,108
5	-0,592	-0,592	-0,440	-0,108	—
$\sum_{i=1}^5 m_i$	-1,299	-0,625	-0,253	+0,684	+1,443
$\sum_{i=1}^5 m_i^2$	0,569	0,246	0,234	0,291	0,680
$n \sum_{i=1}^5 m_i^2$	2,554	1,230	1,121	1,456	3,402
$(\sum_{i=1}^5 m_i)^2$	1,687	0,391	0,064	0,471	2,229
$n \frac{\sum_{i=1}^5 m_i^2}{\sum_{i=1}^5 m_i} - (\sum_{i=1}^5 m_i)^2$	1,142	0,839	1,107	1,488	1,173
$n v_i \cdot \frac{1}{n \sum_{i=1}^5 m_i^2 - (\sum_{i=1}^5 m_i)^2}$	0,57	0,92	1,05	0,94	1,08
$\frac{1}{n v_i}$	0,93	1,09	0,95	1,01	0,92
$\sum_{i=1}^5 \frac{1}{n v_i} = 4,90 \quad n P = \frac{2m}{490} = 2,104$					
$\hat{b}_j = \frac{nP}{n v} - 1$	0,41	1,21	0,93	1,06	0,89

5,0 ✓

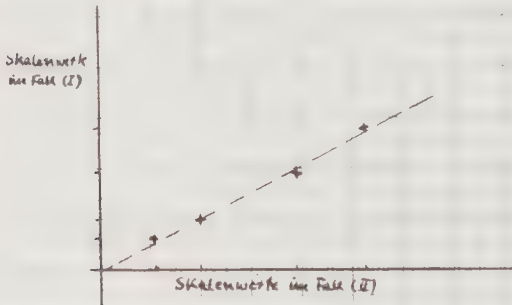
Mit Hilfe von $\hat{b}_j$ können wir nun eine zweite Matrix aufstellen, deren einzelne Werte von $\hat{b}_j \cdot m_{ij}$ gebildet werden.

$\bar{x} + cmis$	1	2	3	4	5	$\sum \hat{b}_j \cdot m_{ij}$
1	—	0,132	0,176	0,437	0,527	1,272
2	-0,098	—	0,050	0,347	0,314	0,613
3	-0,120	-0,065	—	0,056	0,392	0,213
4	-0,325	-0,399	-0,050	—	0,096	-0,728
5	-0,539	-0,431	-0,414	-0,114	—	-1,498
$\sum_{i=1}^5 \hat{b}_j \cdot m_{ij}$	-1,182	-0,763	-0,238	+0,726	+0,129	
$A = \sum_{i=1}^5 \hat{b}_j \cdot m_{ij} - \sum_{i=1}^5 \hat{b}_i \cdot m_{ij}$	-2,454	-1,376	-0,451	-1,454	+2,827	0,0 ✓
$S'_j = \frac{1}{n} (A)$	-0,491	-0,275	-0,090	+0,291	+0,565	
$S_j = S'_j + 0,491$	0,0	0,216	0,401	0,782	1,056	

Über diese Matrix können wir nun nach (64) die einzelnen S_j bestimmen.

Wie wir sehen, sind die Skalenwerte verschieden von denjenigen, die wir bei Fall (I) erhalten haben. Das Schaubild zeigt indessen, daß die Beziehung zwischen beiden linear ist, die Skala wurde also nur gestreckt.

Schaubild 19



Offensichtlich wäre also in unserem Fall die Berechnung nach der zweiten, aufwendigeren Methode nicht notwendig gewesen. Die $\hat{\sigma}_j$ zeigen nur sehr geringe Differenzen, die Annahme $\hat{\sigma}_j$ gleich eins, wäre durchaus zu halten gewesen. Das Beispiel sollte indessen nur zur Illustration des Verfahrens dienen.

d. Als letztes wollen wir noch die Bestimmung der Skalenwerte nach Fall (I) bei unvollständiger Matrix vorführen. Da uns hierfür kein empirisches Material zur Verfügung steht, wollen wir die Daten fingieren.

Die P-Matrix sei:

P-Matrix

Item	1	2	3	4	5
1	-	0,934	0,977	1,000	0,998
2	0,066	-	0,691	0,934	1,000
3	0,023	0,309	-	0,934	1,000
4	0,000	0,066	0,066	-	0,934
5	0,002	0,000	0,000	0,066	-
$\sum p_{ij}$	0,091	1,309	1,734	2,934	3,932

Daraus ergibt sich die M-Matrix. Diejenigen Zellen der M-Matrix, die p-Werten von 0,0 oder 1,0 entsprechen, bleiben leer. Im allgemeinen wird man auch schon bei p-Werten von 0,98 (bzw. 0,02) und darüber keine m-Werte mehr berechnen. Infolge der geringen Überschneidungsfläche der Verteilungen ist die Information hier sehr unscharf.

M-Matrix

Item	1	2	3	4	5
1	0,0	1,5	2,0	-	3,0
2	-1,5	0,0	0,5	1,5	-
3	-2,0	-0,5	0,0	1,5	-
4	-	-1,5	-1,5	0,0	1,5
5	-3,0	-	-	-1,5	0,0
$\sum m_{ij}$	-6,5	-0,5	+1,0	+1,5	+4,5

Nach Gleichung (47) ergibt sich nun folgende Matrix der Differenzen:

Differenz Item	2 - 1	3 - 2	4 - 3	5 - 4
1	1,5	0,5	-	-
2	1,5	0,5	1,0	-
3	1,5	0,5	1,5	-
4	-	0,0	1,5	1,5
5	-	-	-	1,5
$\sum d$	4,5	1,5	4,0	3,0
n_b	3	4	3	2
$\frac{1}{n_b} \sum d$	1,5	0,4	1,3	1,5

Setzen wir nun S_1 gleich null, so erhalten wir folgende Skalenwerte (48):

$$S_1 = \underline{\underline{0,0}}$$

$$S_2 = 0,0 + 1,5 = \underline{\underline{1,5}}$$

$$S_3 = 0,0 + 1,5 + 0,4 = \underline{\underline{1,9}}$$

$$S_4 = 0,0 + 1,5 + 0,4 + 1,3 = \underline{\underline{3,2}}$$

$$S_5 = 0,0 + 1,5 + 0,4 + 1,3 + 1,5 = \underline{\underline{4,7}}$$

Damit beenden wir die Darstellung der Verfahren an Hand praktischer Beispiele. Wir hoffen damit gezeigt zu haben, daß die Konstruktion einer metrischen Skala in praxi nicht so kompliziert ist, wie es zuerst scheinen mag.

C. Tests auf die Richtigkeit der zugrundeliegenden Hypothesen

1. Vorbemerkung

Auf die eben beschriebene Art und Weise haben wir eine metrische Skala für die zu untersuchende Einstellung gewonnen. Ehe wir allerdings die eigentliche "meinungsumfrage", also die Messung der Verteilung der Eigenschaft "Einstellung zu....." auf die Mitglieder der repräsentativen Teil:esamtheit vornehmen können, müssen wir unsere Skala erst noch testen. Die Frage ist: wie gut entspricht unser empirisches Material dem theoretischen Modell? Ist vor allem unsere stets zugrundeliegende Hypothese der Eindimensionalität erfüllt, sind mit:in die Ergebnisse der Beurteilungen konsistent und ist das skalierte Merkmal additiv?

Bei der weiter oben besprochenen Konstruktion einer topologischen Skala konnten wir, infolge des deterministischen Charakters des Modells, den Test nur posteriorisch, nach der eigentlichen Umfrage veranstalten. Wir schlossen aus dem Vorhandensein oder Nichtvorhandensein einer gewissen Systematik in den Antwortkombinationen der einzelnen Befragten, also aus der Konsistenz ikrer Meinungen, auf die Richtigkeit unserer Hypothesen. Additivität zudem forderten wir ja von vornherein nicht, da wir nur eine topologische Skala zu konstruieren beabsichtigten^t.

Hier haben wir es mit einem Wahrscheinlichkeitsmodell zu tun. Das heißt aber, daß wir statistische Signifikanztests auf die Anpassung des empirischen Materials an das theoretische Modell anwenden können. Das wollen wir im folgenden tun.

Wir werden vor allem unsere Skala auf das Vorhandensein von Eindimensionalität testen. Und zwar doppelt: einmal mit Hilfe eines Additivitätstests, zum anderen durch einen Konsistenztest.

Die Abweichungen des empirischen Materials vom theoretischen Modell könnten allerdings auch noch auf das Nichteintreffen anderer Voraussetzungen, als der der Eindimensionalität zurückzuführen sein. Vor allem könnte keine Normalität der Verteilung der Reaktionsprozesse oder (im Fall I) Ungleichheit in den Verteilungen der Differenzen auftreten. Bis zu welchem Grad das Ergebnis unserer Tests von diesen Faktoren beeinflusst wird, kann nicht abgeschätzt werden. Es ist aber anzunehmen, daß er hauptsächlich auf Mängel in der Eindimensionalität reagiert.

2. Ein Additivitätstest

a. Soll unsere Skala wirklich metrischen Charakter haben, so muß sie additive Eigenschaften besitzen. Das heißt z.B., für drei Skalenwerte $S_1 < S_2 < S_3$ müßte

$$(76) \quad S_3 - S_1 = (S_2 - S_1) + (S_3 - S_2)$$

sein. Ist unser Kontinuum nicht eindimensional, so wird sich keine derartige Beziehung ergeben.¹⁾ Wir werden also zuerst eine abgeleitete M-Matrix aufstellen, deren Werte m'_{ij} die

1) Mosteller, F., a.a.O., 1951, S-207-208.

Differenzen zwischen zwei Stimuli auf Grund der errechneten Skalenwerte darstellen.

M'-Matrix

Skalenwerte	S_1	S_2	...	S_j	...	S_n
S_1	-	m'_{12}		m'_{1j}		m'_{1n}
S_2		m'_{21}	-	m'_{2j}		m'_{2n}
						$\vdots$
S_i		m'_{i1}	m'_{i2}	...	m'_{ij}	m'_{in}
						$\vdots$
S_n		m'_{n1}	m'_{n2}	...	m'_{nj}	m'_{nn}

Ausgehend von den m'_{ij} Werten können wir abgeleitete p'_{ij} Werte finden.

P'-Matrix

Skala	1	2	...	j	...	n
1	-	p'_{12}		p'_{1j}		p'_{1n}
2		p'_{21}	-	p'_{2j}		p'_{2n}
						$\vdots$
i		p'_{i1}	p'_{i2}	...	p'_{ij}	p'_{in}
						$\vdots$
n		p'_{n1}	p'_{n2}	...	p'_{nj}	p'_{nn}

Die p'_{ij} Werte sind die Anteile der Schätzungen $j(p)_i$, die unseren Skalenwerten entsprechen müßten. Diese P'-matrix bildet nun unser Ausgangsmaterial für den Test.

b. Die abgeleiteten p'_{ij} Werte werden mit den beobachteten p_{ij} Werten verglichen. Ein einfaches Maß für die Anpassung des empirischen Materials an das theoretische Modell, wäre dann schon die durchschnittliche absolute Abweichung der p'_{ij} Werte von den p_{ij} Werten

$$(77) \quad \overline{p'_{ij} - p_{ij}} = \frac{1}{n(n-1)} \sum |p'_{ij} - p_{ij}|$$

Ist diese Abweichung "klein", so erscheint die Hypothese als gesichert. Dieses Maß ist jedoch relativ grob. Vor allem kann man es nur anwenden, wenn man erhebliche Erfahrung mit Befragungen dieser Art hat.

Als wesentlich exakteres statistisches Maß bietet sich an. Der χ^2 Test läßt sich hier relativ leicht durchführen. Wir testen dabei ganz allgemein die O-Hypothese, daß die Annahmen des Modells zu halten sind. Die Alternative zur O-Hypothese ist ebenso allgemein, sie sagt praktisch nur, daß irgendwo ein Fehler stecken muß. Wir werden bei einem Verwerfen der O-Hypothese natürlich zuerst auf einen Fehler in der Additivität, also ein Fehlen der Eindimensionalität schließen. Das Nichteintreffen der O-Hypothese kann allerdings auch darauf zurückzuführen sein, daß z.B. die Verteilungen der Reaktionsprozesse auf dem Kontinuum nicht normal oder die Standardabweichungen der Differenzen nicht gleich sind. Theoretisch kann also der Grund für eine Zurückweisung der O-Hypothese in jedem einzelnen oder einer Kombination dieser Punkte zu suchen sein. Bei der praktischen Durchführung werden wir indessen haupt-

sächlich Fehler der Additivität für ein Verwerfen der O-Hypo-
these verantwortlich machen.

Zur Berechnung von χ^2 werden wir uns einer Vereinfachung bedienen. Wir werden sowohl die abgeleiteten wie die beobachteten Werte p'_{ij} und p_{ij} nach der Formel

$$(78) \quad \theta = \arcsin \sqrt{\rho}$$

transformieren. Die Transformation ist an Hand einer Tabelle von Snedecor/Bliss leicht durchzuführen.¹⁾ θ ist annähernd normal verteilt mit einem σ_θ^2 von 821/N

$$(79) \quad \sigma_\theta^2 = \frac{821}{N}$$

wobei N in unserem Fall die Zahl der Beurteiler ist. Den Abweichungen $p'_{ij} - p_{ij}$ entsprechen jetzt $\theta'_{ij} - \theta_{ij}$. χ^2 ergibt sich dann als

$$(80) \quad \chi^2 = \frac{N \sum (\theta'_{ij} - \theta_{ij})^2}{821}$$

Die Zahl der Freiheitsgrade ist dabei offensichtlich

$$(81) \quad F_g^I = \frac{(n-1) \cdot (n-2)}{2}$$

im Fall (I) und

$$(82) \quad F_g^I = \frac{n(n-1)}{2} - 2n - 1$$

1) Snedecor, G.W.: Statistical Methods (5. Aufl.) Ames, Iowa, 1956. Die $\arcsin$ Werte selbst wurden von Bliss berechnet.

im Fall (II), da hier die G_j mit einbezogen werden müssen.
 Als signifikant wollen wir diejenigen Werte von χ^2 betrachten,
 die einer Wahrscheinlichkeit von 0,05 oder weniger entsprechen.

c. Beispielshalber wollen wir den χ^2 Wert für unsere oben
 in Fall (I) erhaltene Skala berechnen.

Wir können folgende M'-Matrix aufstellen:

S	0	0,125	0,1999	0,357	0,549
0	-	0,125	0,1999	0,357	0,549
0,125	-0,125	-	0,074	0,100	0,424
0,1999	-0,1999	-0,074	-	0,188	0,350
0,357	-0,357	-0,100	-0,188	-	0,160
0,549	-0,549	-0,424	-0,350	-0,160	-

Davon ausgehend kommen wir zu der abgeleiteten P'-Matrix, der
 die beobachtete P-Matrix gegenübersteht.

P'-Matrix

xi	1	2	3	4	5
1	-	0,550	0,579	0,671	0,709
2		-	0,530	0,584	0,664
3			-	0,574	0,637
4				-	0,564
5					-

P-Matrix

xi	1	2	3	4	5
1	-	0,543	0,574	0,660	0,729
2		-	0,521	0,628	0,638
3			-	0,521	0,670
4				-	0,590
5					-

Als Summe der Abweichungen würden wir 0,502 erhalten, damit ergäbe sich

$$\overline{p'_{ij} - p_{ij}} = \frac{1}{n(n-1)} \sum |p'_{ij} - p_{ij}| = 0,025$$

Dieser Wert ist relativ gering. Der "primitive" Test deutet also die Richtigkeit unserer Hypothesen an.

Wollen wir nun mittels χ^2 testen, so müssen wir die P' und die P-Matrix zuerst in eine Θ' bzw. Θ Matrix umwandeln.

Θ' -Matrix

St	1	2	3	4	5
1	—	47,87	49,54	53,79	57,75
2		—	46,72	46,68	59,57
3			—	49,26	52,95
4				—	48,68
5					—

Θ -Matrix

St	1	2	3	4	5
1	—	47,47	49,26	54,33	58,24
2		—	46,20	52,42	59,01
3			—	46,20	54,94
4				—	47,47
5					—

Als Matrix der Abweichungen ($\Theta' - \Theta$) ergibt sich dann

$\Theta - \Theta'$ -Matrix

St	1	2	3	4	5
1	—	-0,40	-0,28	+0,54	+0,89
2		—	-0,52	+3,74	-1,56
3			—	-3,06	+4,99
4				—	-1,21
5					—

Damit für χ^2

$$\chi^2 = \frac{N(\sum (o-e)^2)}{821} = \frac{32,839}{\frac{821}{N}}$$

N, die Zahl der Beurteiler muß, wie wir der weiter oben aufgeführten H-Matrix entnehmen, 94 gewesen sein (da $h_{ij} + h_{ij} = 94$). Es ergibt sich also

$$\chi^2 = 3,76$$

Für sechs Freiheitsgrade können wir aus der Tabelle eine Wahrscheinlichkeit von rund 70 % entnehmen.

Wenn wir denjenigen Wert von χ^2 als Signifikanzschwelle betrachten, dem eine Wahrscheinlichkeit von 5% oder weniger entspricht, so müßte in unserem Fall $\chi^2 = 12,592$ oder größer sein um anzuzeigen, daß unsere Annahmen falsch sind. So aber können wir schließen, daß unsere Skala wirklich additiv, also die Voraussetzung eines zugrundeliegenden eindimensionalen Kontinuum richtig ist.

3. Ein Konsistenztest

a. Ein ausgezeichnete Test auf die Urteilsfähigkeit der einzelnen Beurteiler, wie auch auf unsere Annahme eines eindimensionalen Kontinuum ist der Konsistenztest. Allerdings erfordert er einen erheblichen Rechenaufwand, da man bei der Bestimmung der Konsistenzkoeffizienten auf die einzelnen Fragebögen zurückgehen muß.

Vorausgesetzt, die einzelnen Stimuli wären klar auf einem eindimensionalen Kontinuum angeordnet und die Beurteiler

würden ihre Aufgabe sorgfältig und mit gutem Urteilsvermögen durchführen, so müßte praktisch die Antwortmatrix für jeden einzelnen Beurteiler wie folgt aussehen:

Individualmatrix (Idealfall)

	1	2	3	4	5	6	7	8	9
1	-	1	1	1	1	1	1	1	1
2	0	-	1	1	1	1	1	1	1
3	0	0	-	1	1	1	1	1	1
4	0	0	0	-	1	1	1	1	1
5	0	0	0	0	-	1	1	1	1
6	0	0	0	0	0	-	1	1	1
7	0	0	0	0	0	0	-	1	1
8	0	0	0	0	0	0	0	-	1
9	0	0	0	0	0	0	0	0	-

Dabei bedeutet eine 1 das Stimulus j für positiver oder allgemein ranghöher erachtet wurde als Stimulus i, eine 0 umgekehrt.

Soweit der Idealfall. Nun ist es aber durchaus möglich, daß sogenannte Triaden auftreten. Das heißt, der einzelne Beurteiler hat zwar von drei rangmäßig angeordneten Stimuli i, j, k, Stimulus i für positiver als Stimulus j und Stimulus j für positiver als Stimulus k beurteilt, dann aber Stimulus k für positiver als Stimulus i bezeichnet.

Also

$$i > j > k$$

aber

$$k > i$$

Beurteilungen dieser Art verraten offensichtlich Inkonsistenz. Sie werden in der Individualmatrix wie folgt erscheinen (fiktives Beispiel für 9 Stimuli):

Individualmatrix

Stim.	1	2	3	4	5	6	7	8	9
1	-	1		1	1	1	1		
2		-	1	1		1		1	1
3	1		-			1			1
4				-		1	1		
5				1	-	1		1	
6						-	1	1	1
7							-		1
8	1		1			1		-	
9					1				-
Σ	4	4	4	4	4	4	6	4	6
Σ^2	4	4	16	16	16	16	36	16	36

$$\Sigma = 160$$

Der Grund für das Auftauchen derartiger Triaden kann, wie schon gesagt, zweifach sein:

einmal kann die zugrundeliegende Hypothese der Eindimensionalität falsch sein, dann werden gleichzeitig bei allen Beurteilern eine große Zahl von Triaden auftauchen,

zweitens kann der Fehler beim Beurteiler selbst liegen, er war nicht sorgfältig genug oder ist ganz einfach nicht urteilsfähig. In diesem Fall wird sich bei einzelnen Beurteilern eine

große Zahl von Triaden ergeben, während sie im allgemeinen gering sind.

Wir können nun eine Maßzahl entwickeln, die in bestimmten Grenzen die Konsistenz der einzelnen Beurteilungen angibt. Ferner läßt sich testen, ob zwischen den empirischen Beurteilungen und solchen, die theoretisch nach dem Zufallsprinzip getroffen wurden, ein wesentlicher Unterschied besteht.

b. Die erste Maßzahl läßt sich leicht berechnen. Das Problem ist uns nicht neu, Kendall hat es bereits ausführlich behandelt.¹⁾

Wenn die Zahl der zu beurteilenden Stimuli ungerade ist, können sich als Maximum $(n^3-n)24$ Triaden ergeben. (Wobei n die Zahl der Stimuli bezeichnet). Ist die Zahl der Stimuli gerade, so beträgt das Maximum der Triaden $(n^3-4n)24$. Als Konsistenzmaß K können wir daher schreiben:

$$(83) \quad K^u = 1 - \frac{24T}{n^3-n}$$

für ungerade n und

$$(84) \quad K^g = 1 - \frac{24T}{n^3-4n}$$

für gerade n .

T ist dabei die jeweils beobachtete Zahl der Triaden, sie ergibt sich aus der Individualmatrix als

1) Kendall, M.G.: Rank Correlation Methods, London, 1948.

$$(85) \quad T = \frac{1}{12} \left[n(n-1)(2n-1) - 6 \sum a^2 \right]$$

wobei a die Zahl der Eintragungen in einer Spalte ist.

K wird nun zwischen 0 und 1 schwanken. Sind in den Angaben eines Beurteilers überhaupt keine Triaden vorhanden ($T = 0$), so wird sich 1 ergeben, sind alle nur möglichen Triaden in seine Matrix, so wird $K = 0$.

Über den Grad der Konsistenz oder Inkonsistenz hinaus können wir aber auch noch die Hypothese testen, ob sich bei rein zufälliger Abgabe der Beurteilungen nicht die gleiche Zahl von Triaden ergeben hätte. Wir haben damit wieder ein typisches Test-Problem.

Für Werte von $n < 7$ liegen Tabellen für die Wahrscheinlichkeiten völlig zufällig getroffener Beurteilungen vor. Sie wurden von Kendall in einem ähnlichen Zusammenhang berechnet.¹⁾ Für $n \geq 7$ ergibt sich das χ^2 aus der Formel

$$(86) \quad \chi^2 = \frac{8}{n-4} \left(\frac{1}{4} \frac{n!}{3!(n-3)!} - T + \frac{1}{2} \right) + \bar{f}_g$$

Die Zahl der Freiheitsgrade ist dabei

$$(87) \quad \bar{f}_g = \frac{n(n-1)(n-2)}{(n-4)^2}$$

Wir können uns die Ableitung der Formel sparen, sie findet sich bei Kendall, der sich eingehend mit einem ähnlichen Problem befaßt hat.

1) Kendall, M.G.: Rank Correlation Methods, London, 1948.

c. Ein echtes praktisches Beispiel können wir für den Konsistenztest nicht geben, da uns die dem vorigen Beispiel zugrundeliegenden Einzelurteile nicht zur Verfügung stehen. Wir können indessen K und χ^2 für unsere oben gezeigte inkonsistente Individualmatrix berechnen, um so wenigstens eine Andeutung der Methode zu geben. Es wird sich dort ergeben:

$$T = 22 \quad \text{für } n = 9 \text{ und } a^2 = 160$$

das heißt

$$K = 0,238$$

und

$$\chi^2 = 19,36 \quad \text{bei } F_g = 20,16$$

wofür

$$p = 0,50$$

Die Wahrscheinlichkeit dafür, einen T-Wert von 22 dann zu erhalten, wenn die Beurteilungen völlig zufällig erfolgt waren, ist 50%. Ein gewisser Grad von Konsistenz dürfte also in den vorliegenden Beurteilungen vorhanden sein, wenn sie auch weit vom Idealfall entfernt sind.

4. Zusammenfassung

Statistische Tests auf die Richtigkeit der zugrundeliegenden Hypothesen lassen sich mithin durchführen. Wir können, mit Einschränkungen, nachprüfen, ob die zwei Hauptanforderungen,

die wir an unsere metrische Skala stellen müssen: Konsistenz und Additivität, erfüllt sind.

Wir sagten eben "mit Einschränkung" - denn wir müssen allerdings gestehen, daß beide der vorgeschlagenen Tests nicht eindeutig auf das Vorhandensein oder Nichtvorhandensein dieser Voraussetzungen im einzelnen hinweisen können. Wir sind jeweils nur in der Lage, quasi ein ganzes Bündel von Hypothesen auf einmal zu prüfen. Im Fall des Additivitätstests ist es die Hypothese der Additivität (also der Eindimensionalität des Kontinuum) plus der Hypothese der normalen Verteilung der Reaktionsprozesse plus der Hypothese konstanter Streuungen der Differenzen. Im Fall des Konsistenztests wieder die Hypothese der Eindimensionalität des Kontinuum plus der Hypothese einer guten Urteilsfähigkeit der einzelnen Beurteiler.

Allerdings ist bei der praktischen Durchführung der Messung die Lage weniger schwierig, als es hier scheinen mag. Im ersten Fall werden wir annehmen können, daß der Test am stärksten auf Mangel in der Eindimensionalität reagiert. Im zweiten Fall wird man aus der Verteilung der K und χ^2 für die einzelnen Beurteilungen schließen können, ob der Fehler eher bei der Voraussetzung der Eindimensionalität oder bei der Urteilsfähigkeit einzelner Beurteiler liegen wird. Ein Kollektivmaß von der Form

$$(87) \quad K = \frac{\sum_{i=1}^N |K_i - \bar{K}|}{N}$$

gibt darüber Aufschluß. K_i seien die Konsistenzkoeffizienten

der einzelnen Beurteiler, $\bar{K}$ deren Mittelwert. Aus der relativen Größe von K und $\bar{K}$ könnte man dann etwa folgende Schlüsse ziehen:

$K \backslash \bar{K}$	hoch	niedrig
hoch	Eindimensionalität vermutet, aber einige schlechte Beurteiler vorhanden	Weder Eindimensionalität noch Urteilsfähigkeit vermutet
niedrig	Sowohl Eindimensionalität wie Urteilsfähigkeit sicher	Eindimensionalität fehlt, über Urteilsfähigkeit läßt sich nichts sagen. Oder generell schlechte Urteilsfähigkeit

Mit Hilfe dieser zwei Tests und Intuition und Sachkenntnis kann man also relativ sicher feststellen, ob die der Konstruktion unserer metrischen Skala zugrundeliegenden Hypothesen richtig sind.

Damit aber ist der Bau des Maßstabs abgeschlossen und wir können jetzt an die eigentliche "meinungsumfrage", die Messung der Verteilung der interessierenden Eigenschaft auf die repräsentative Teilgesamtheit ihrer Träger beginnen.

D. Die Messung der Einstellung

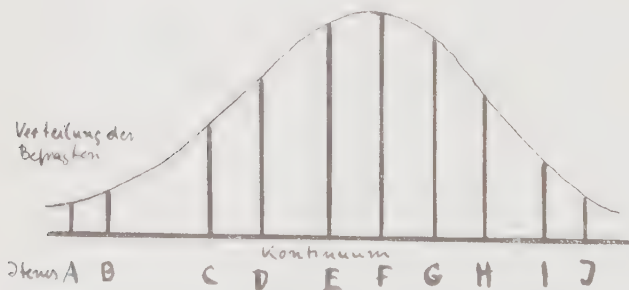
1. Die Skalierung der Befragten

Haben wir erst einmal eine metrische Skala für die zu untersuchende Einstellung konstruiert, so ist die eigentliche

Messung der Verteilung dieser Einstellung auf die Befragten trivial. Sie birgt keine größeren Probleme in sich als etwa das Messen und Wiegen einer Gruppe von Kindern bei der Reihenuntersuchung einer Schulklasse.

Handelt es sich bei unseren Stimuli um Punktitems (siehe Seite 23), so könnten wir die Befragten etwa anweisen, auf einem Fragebogen dasjenige anzukreuzen, das ihrer Einstellung am besten entspricht. Wir erhalten dann für jedes Item bestimmte Antworthäufigkeiten und können die Verteilung annähern. Das Ergebnis wird etwa so aussehen:

Schaubild 20



Eine weitere Möglichkeit wäre, die Befragten zu bitten, die drei Items anzukreuzen, die ihrer Einstellung am nächsten sind. Der Mittelwert dieser drei Items würde uns dann für jeden Befragten einen bestimmten Skalenwert liefern. Allerdings wird dieser Skalenwert nicht besonders exakt sein.

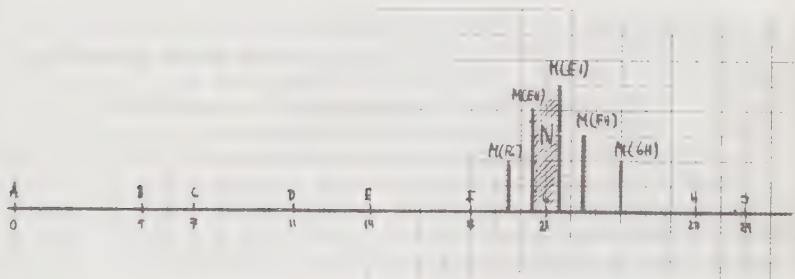
Weit vielversprechender erscheint uns folgende Methode: Die Befragten werden angewiesen, aus den vorgelegten Items oder Stimuli die drei oder vier herauszusuchen, die ihrer

Position auf dem Kontinuum am besten entsprechen und diese nach ihrer individuellen Präferenz zu ordnen. Also ein Vorgehen ähnlich dem in Kapitel 2 unter der Überschrift "Komparative Modelle" besprochenen. Dadurch kann eine viel feinere Bestimmung der Verteilung erreicht werden.

Nehmen wir an, wir hätten neun Items A,B,.....,I mit folgenden Skalenwerten:

A=0, B=5, C=7, D=11, E=14, F=18, G=20, H=27, I=29.

Schaubild 21



Die Position jedes Befragten könnte nun an Hand der individuellen Präferenzskala recht genau festgelegt werden. Nehmen wir an, er hätte die Anordnung GFHE getroffen, so können wir daraus folgende Information schöpfen:

- | | | |
|----------------------|---------------------|-------------------|
| 1.) G(1) | $M(FG) < N < M(GH)$ | $19,5 < N < 24,0$ |
| 2.) G(1)F(2) | $M(FG) < N < M(FH)$ | $19,5 < N < 22,5$ |
| 3.) G(1)F(2)H(3) | $M(EH) < N < M(FH)$ | $20,0 < N < 22,5$ |
| 4.) G(1)F(2)H(3)E(4) | $M(EH) < N < M(EI)$ | $20,5 < N < 21,5$ |

$M(FG)$ bezeichnet dabei z.B. den Mittelpunkt zwischen F und G.

N ist die Position des Befragten auf dem Kontinuum. Auf diese Art und Weise haben wir das Intervall, in dem sich die Position

des Befragten auf dem Kontinuum befinden kann von 4,5 Einheiten (für den Fall, daß er nur Item G angekreuzt hatte, siehe oben) auf eine Einheit begrenzt. Der Informationsgewinn ist also erheblich. Die Intervallgrenzen für die möglichen Präferenzkombinationen lassen sich vorner berechnen, so daß bei der eigentlichen Befragung nur die auf die einzelnen Kombinationen entfallenden Häufigkeiten abgezählt zu werden brauchten. In unserem Fall ergäben sich ca. 30 Befragtengruppen. Die Intervallgrenzen waren jeweils

ABCD		-	2,5
BACD	2,5	-	4,5
BCAD	4,5	-	5,5
BCDA	5,5	-	6,5
CBDA	6,5	-	7,5
CBDE	7,5	-	8,5
CDBE	8,5	-	9,5
DCBE	9,5	-	9,5
DCEB	9,5	-	10,5
DECB	10,5	-	11,5
DECF	11,5	-	12,5
.	.	.	.
.	.	.	.
.	.	.	.
IHGF	25,5	-	

Wir besitzen damit die Häufigkeiten für eine große Zahl von Befragtengruppen, die das Kontinuum relativ eng abteilen. Die Verteilung der Einstellung auf die Gesamtheit läßt sich daraus sehr gut abschätzen.

2. Die Bestimmung des O-Punkts

a. Wir haben jetzt die Verteilung der untersuchten Eigenschaft auf die Befragten relativ exakt festgestellt. Was wir

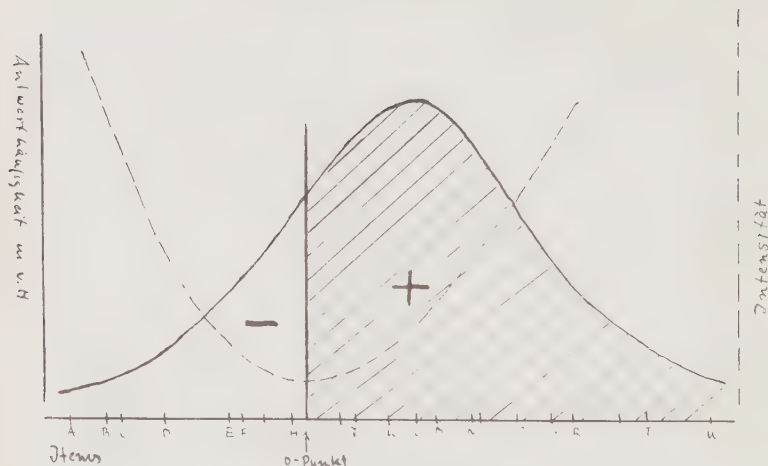
nun eigentlich "dafür" und wer "dagegen" ist. Es ergibt sich also wieder das Problem des rationalen O-Punktes.

Bei der Konstruktion der metrischen Skala hatten wir ihn willkürlich festgesetzt. Das durften wir dort ohne weiteres, Jetzt aber müssen wir den O-Punkt so verschieben, daß er eine Trennung für die wahren Anteile des "Pro" und "Kontra" bildet. Das Problem ist uns nicht neu, wir haben es bei der topologischen Messung bereits kennengelernt.

Genauso wie dort könnten wir es nun auch hier mit Hilfe einer Intensitätsskala lösen. Den einzelnen Items oder Stimuli werden drei, fünf oder mehr Antwortmöglichkeiten zugeordnet (zum Beispiel: "stark dafür, dafür, etwas dafür, etwas dagegen, dagegen, stark dagegen") und die Befragten angewiesen, durch sie die Intensität ihrer positiven oder negativen Antwort zu kennzeichnen. Die Antworten auf diese Fragen werden nun einer einfachen Skalogrammanalyse unterzogen. Ein Mittelwert der Positionen derjenigen Befragten, die in dieser Skalogrammanalyse die niedrigsten Rangziffern erhalten haben, stellt dann den rationalen O-Punkt dar. (Näheres dazu findet sich auf Seite 53f.).

Das Ergebnis der Einstellungsmessung plus der Intensitätsanalyse sieht etwa wie folgt aus:

Schaubild 22



b. Eine andere, zwar elegante, aber nicht immer anwendbare Methode erlaubt schon bei der Bestimmung des Skalenwerts der einzelnen Items diesen auf einen rationalen 0-Punkt zu beziehen.

Angenommen, wir hätten zwei Stimuli, von denen wir glauben, daß sich das eine auf der positiven, das andere auf der negativen Seite des Kontinuums befindet. Nun könnte man von den Beurteilern Vergleiche der Form: "Stimulus j ist positiver als Stimulus i negativ ist" erhalten. Die Beurteiler müßten also praktisch die Summe der Skalenwerte beider Stimuli bilden und, wenn diese positiv ist, mit "ja", wenn sie negativ ist, mit "nein" antworten. (Die Fragen müßten etwa lauten: "Würden Sie Stimulus i nehmen, wenn Sie Stimulus j dazu bekämen?")

Diese Schätzung "j ist positiver als i negativ" bezieht sich dann aber auf einen rationalen O-Punkt.¹⁾

Kombiniert man die Ergebnisse dieses Vergleichs mit den üblichen Beurteilungen der Stimuli, wie sie weiter oben beschrieben wurden, so kann man nach S_j^r - den auf einen rationalen O-Punkt bezogenen Skalenwerten - auflösen:

$$(18) \quad S_j^r - S_i^r = m_{ij} \sqrt{\sigma_j^2 + \sigma_i^2 - 2r_{ij} \sigma_i \sigma_j}$$

Analog können wir jetzt erhalten

$$(19) \quad S_j^+ + S_i^+ = m_{ij}^+ \sqrt{\sigma_j^2 + \sigma_i^2 - 2r_{ij}^+ \sigma_i \sigma_j}$$

wobei m_{ij}^r den Anteilen entspricht, die wir durch die (+)(-)Vergleiche erhalten haben. Nach den Vereinfachungen für den Fall I ergäbe sich dann

$$(89) \quad S_j^r - S_i^r = m_{ij}$$

und

$$(90) \quad S_j^+ + S_i^+ = m_{ij}^+$$

Daraus, nach Addition

$$(91) \quad 2 S_j^r = m_{ij} + m_{ij}^+$$

1) siehe dazu Horst, A.P.: A method for determining the absolute value of a series of stimulus situations. J.Educ.Psych., 23, 1932, S.418-440; zitiert nach W.S. Torgerson, a.a.O., 1958, S.197.

Die zu minimierenden Abweichungen A^r wären

$$(92) \quad A^r = \sum_{i=1}^n \sum_{\substack{j=1 \\ i \neq j}}^n (m_{ij} + m_{ij}^r - 2S_j^r)^2$$

das heißt

$$(93) \quad \frac{1}{4} \frac{dA^r}{dS_j^r} = - \sum_{i=1}^n (2S_j^r - m_{ij} - m_{ij}^r) = 0 \quad (j = 1, 2, \dots, n)$$

und

$$(94) \quad 2(n-1)S_j^r = \sum_{i=1}^n (m_{ij} + m_{ij}^r) \quad (j = 1, 2, \dots, n)$$

daraus ergibt sich

$$(95) \quad S_j^r = \frac{1}{2(n-1)} \sum_{i=1}^n (m_{ij} + m_{ij}^r) \quad (j = 1, 2, \dots, n)$$

Die S_j^r beziehen sich jetzt auf einen rationalen O-Punkt. Werden die auf diese Art und Weise bewerteten Stimuli bei einer Befragung verwendet, so würden wir sofort die Anteile der zu dem Untersuchungsobjekt positiv bzw. negativ eingestellten Befragten erhalten.

Allerdings wird sich diese Methode nur selten anwenden lassen. Beim Warentest vielleicht, bei Items ist indessen der (+)(-)Vergleich kaum denkbar. Wir werden also meist auf die nicht ganz so elegante, aber dafür immer anwendbare Methode des Intensitätsvergleichs zurückgreifen müssen.

3. Zusammenfassung

Damit ist auch die metrische Skalierung der Befragten abgeschlossen. In den vorigen Abschnitten hatten wir die Methoden aufgezeigt, mittels derer ein Maßstab für eine bestimmte, bislang qualitative Eigenschaft konstruiert werden kann. Jetzt haben wir diesen Maßstab an eine repräsentative Teilgesamtheit angelegt und die Verteilung der betreffenden Eigenschaft auf die Befragten gemessen.

Je nach der Aufgabe, die wir dabei den Befragten stellen, sind drei Methoden dafür denkbar:

- a. Jeder Befragte kennzeichnet denjenigen Stimulus oder dasjenige Item, das seiner Einstellung am besten entspricht. Dadurch werden die Befragten in ebensoviele Gruppen wie Items oder Stimuli existieren aufgeteilt. Die Häufigkeiten für diese Gruppen geben uns dann ein Bild der zugrundeliegenden Verteilung der Eigenschaft.
- b. Jeder Befragte sucht aus den Items oder Stimuli die drei oder vier heraus, die seiner Einstellung am ehesten entsprechen. Sein individueller Skalenwert ergibt sich dann aus dem Durchschnitt der Skalenwerte der angegebenen Items oder Stimuli. Diese Methode wird allerdings etwas verzerrte Ergebnisse liefern, außerdem täuscht sie dadurch, daß sie jedem Befragten einen exakten Skalenwert zuordnet eine falsche Genauigkeit vor.
- c. Wie vorhin sucht wieder jeder Befragte drei oder vier Items oder Stimuli aus, ordnet sie jetzt aber nach seinen individuellen Präferenzen. Dadurch können die Befragten

in eine sehr große Zahl von Befragtengruppen, die jede eine ganz bestimmte individuelle Präferenzskala besitzen, aufgeteilt werden. Somit können die Befragten weit genauer auf dem Kontinuum verteilt und damit eine exaktere Messung erreicht werden. Durch diese Methode wird wohl der größte Informationsgewinn erzielt.

Das zweite Problem, das sich uns bei der Messung stellte, war die Bestimmung eines rationalen O-Punkts. Hier haben wir zwei Methoden gezeigt, die seiner Lösung dienen können:

- a. Wie bei der topologischen Messung kann die Trennung der zu dem Untersuchungsobjekt generell positiv stehenden Befragtengruppen von den zu ihm negativ stehenden durch eine Intensitätsanalyse erreicht werden. Das heißt, der rationale O-Punkt wird durch das Mittel der Skalenwerte jener Items oder Stimuli gekennzeichnet, die von den Befragten mit der indifferentesten Einstellung gekennzeichnet worden waren.
- b. Bereits bei der Bestimmung des Skalenwerts der Items oder Stimuli kann man diesen unter Umständen auf einen rationalen O-Punkt beziehen. Das dazu vorgeschlagene Verfahren einer quasi summativen Schätzung ist allerdings nur in wenigen Fällen anwendbar.

Mit der metrischen Messung der Verteilung der Einstellung auf die Befragten ist auch der letzte Baustein unseres Modells gesetzt.

Damit ist allerdings das Problem der Messung noch lange nicht völlig gelöst. Vor allem interessiert die Frage: Was geschieht, wenn wir die Hypothese eines eindimensionalen Kontinuum fallen

lassen? Sind mehrdimensionale Modelle möglich und denkbar?
Zu dieser Frage wollen wir nach einer kurzen Darstellung unserer Gesamtergebnisse im folgenden Kapitel Stellung nehmen.

4. Kapitel: Ergebnisse und Ausblick

=====

1. Ergebnisse

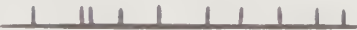
Wir sind jetzt in der Lage, eine abschließende Gesamtdarstellung der Voraussetzungen, Probleme und Ergebnisse unserer Modellkonstruktion zu geben.

a. Um eindimensionale Meinungen und Einstellungen, oder im allgemeinen qualitative, psychische Eigenschaften fundamental zu messen, gehen wir von folgenden Voraussetzungen aus:

1. Jeder Träger der Eigenschaft besitzt diese in einem - zumindest zu einem bestimmten Zeitpunkt - konstanten Ausmaß.
Das heißt, er hat eine feste Position auf dem eindimensionalen Kontinuum der Eigenschaft.

Befragte

Kontinuum

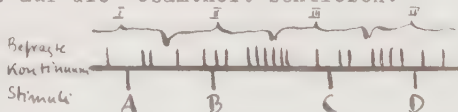


2. Genauso sind aber auch eine Reihe von Stimuli (verbale, in Form von Feststellungen oder Behauptungen, kurz Items - oder dingliche, Warenproben etc.) denkbar, die ebenfalls

die betreffende Eigenschaft in feststellbaren bestimmten Ausmaß besitzen und ebenfalls eine genaue Position auf dem Kontinuum haben.



3. Wenn es nun gelänge, die Position der Stimuli auf dem Kontinuum entweder rangmäßig oder gar metrisch zu bestimmen, so könnten wir diese wiederum als Maßstab für die Position der Befragten auf dem Kontinuum benutzen. Jeder Befragte müßte dann in irgendeiner Weise kenntlich machen, mit welchen Stimuli er übereinstimmt, bzw. welche seiner eigenen Einstellung am ehesten entsprechen. Aus diesen Antworten der Befragten könnten wir dann auf die Verteilung der Eigenschaft auf die Gesamtheit schließen.



Voraussetzung dafür ist natürlich, daß erstens die Skalierung der Stimuli, mithin die Konstruktion des Maßstabes eindeutig erfolgen kann und zweitens beim Anlegen dieses Maßstabes an die Befragten deren Reaktion nur eine Funktion der von den Stimuli in verschiedenem Ausmaß beherbergten Eigenschaft sein wird.

b. Mithin stellen sich drei Probleme:

1. Methoden zu finden, die es uns erlauben, aus der großen Zahl der möglichen Items (oder Stimuli) diejenigen herauszufinden,

die am geeignetsten für diese Aufgabe sind und diese topologisch oder metrisch zu skalieren. Also das Problem der Konstruktion des Maßstabes.

2. Weiterhin Methoden zu finden, die nun der eigentlichen Messung der Einstellung, also dem Anlegen des Maßstabes an die Befragten dienen können. Diese Methoden müßten garantieren, daß mit Hilfe der Stimuli die Position der Befragten auf dem Kontinuum sicher und möglichst scharf gefunden werden kann.
3. schließlich Tests zu entwickeln, die uns zeigen, ob unsere Skalierung überhaupt sinnvoll war, also ob die allem zugrundeliegende Hypothese eines eindimensionalen Kontinuum erfüllt ist. Ist sie es nicht, so ist sowohl die Anordnung der Stimuli wie die der Befragten auf dem Kontinuum nicht invariant, eine Messung in der hier vorgeschlagenen Form wird also sinnlose Ergebnisse liefern.

c. Diese Probleme können wir praktisch in zwei mal zwei Schritten lösen.

Skalierung	Stimuli (Items)	Befragten
topologisch	1	2
metrisch	3	4

Wir haben zuerst Methoden der Auswahl und rangmäßigen Skalierung der Items und Stimuli aufgezeigt (1). Dann haben wir den so entwickelten Maßstab den Befragten vorgelegt und die Verteilung

Schaubild 23

Konstruktion des Maßstabes
(Skalierung der Items oder Stimuli)

Messung der Verteilung
(Skalierung der Befragten)

durch

eine bewußt ausgewählte Beurteilergruppe

eine zufällig ausgewählte repräsentative Teilgesamtheit der Einstellungsträger

topologische
messung

(1)

Aufgabe:
Auswahl der geeignetsten Items, deren top. Anordng.
Methode:
Berechnung der Kriterien α_1, α_2 und α_3 .
Ergebnis:

Kontinuum
A > B > C > D ...
Stimuli

(2)

Aufgabe:
Bestimmung der Rangordnung von Befragten (und Items)
Test auf Eindimensionalität
Methode:
Kategorische oder komparative messung.
Ergebnis:



metrische
messung

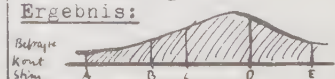
(3)

Aufgabe:
Bestimmung des Skalenwerts der Items oder Stimuli
Test auf Hypothesen
Methode:
Paarweiser Vergleich, Additivitäts- und Konsistenztest.
Ergebnis:



(4)

Aufgabe:
Bestimmung des Skalenwerts der Befragten, damit Messung der Verteilung.
Methode:
Kategorische oder komparative Messung.
Ergebnis:



der Einstellung auf sie topologisch gemessen (2). Ein Test auf das Zutreffen unserer Hypothese liefert uns hier das Vorhandensein einer bestimmten Systematik in den Antwortkombinationen. Soweit die topologische Messung.

Daraufhin sind wir zu der metrischen Messung übergegangen. Das Problem der Bestimmung des Skalenwerts der Stimuli (3) ließ sich lösen, ohne daß von den Beurteilern mehr als ein rangmäßiger Vergleich zweier Stimuli gefordert wurde. Dieser metrische Maßstab wurde nun wieder zu der - jetzt ebenfalls metrischen - Messung der Einstellung der Befragten verwendet (4). Der Test auf die zugrundeliegenden Hypothesen konnte hier schon auf Stufe (3) durchgeführt werden. In Schaubild 23 ist nun das Gesamtmodell schematisch dargestellt. Wir sehen, wie von Stufe zu Stufe, die erzielte Information reichhaltiger wird. Auf der letzten Stufe haben wir - ein Zutreffen aller unserer Hypothesen vorausgesetzt - eine exakte metrische Messung der qualitativen, besser bis dahin qualitativen, Eigenschaft erreicht. Diese qualitative Ranginformation ist aber erzielte weder über einen Index noch sonst ein abgeleitetes Maß. Wichtig ist dabei, daß die metrische Information erzielt wurde, ohne daß von den Beurteilern jemals mehr als rangmäßige Vergleiche der einzelnen Stimuli oder Items gefordert wurden. Wir haben also niemals eine direkte Schätzung vornehmen lassen.

Bei der praktischen Anwendung des Modells für die Messung einer Einstellung wird man nicht immer bis zu der perfekten metrischen Skalierung zu gehen brauchen. Je nach dem Ziel der Befragung wird man sich auch mit gröberen Skalierungen zufrieden geben können. Wenn zum Beispiel nur ein möglichst unver-

zerrter Schätzwert für den Anteil derjenigen Befragten, die "für" ein bestimmtes Untersuchungsobjekt sind, zu dem der "dagegen" ist angestrebt wird, so mag eine einfache Skalogramm-Analyse grob ausgewählter Items, plus einer Intensitäts-Skala genügen. Zur exakten - und natürlich auch aufwendigeren - metrischen Skalierung wird man nur dann übergehen, wenn es wirklich auf eine genaue Kenntnis der Verteilung der betreffenden Eigenschaft ankommt. Bei vielen Problemen der Marktforschung ist dies der Fall.

2. Ausblick

a. Unser Modell steht und fällt mit der Hypothese der Eindimensionalität. Sobald die untersuchte Eigenschaft kein eindimensionales Kontinuum besitzt, werden unsere Methoden sinnlose Ergebnisse liefern. Sowohl die Rangordnung der Stimuli wie die der Befragten ist dann nicht mehr invariant.

Nun hätten wir allerdings gesagt, daß sich die meisten mehrdimensionalen Eigenschaften durch unzureichende Spezifikation in eindimensionale umwandeln lassen. Das wird indessen nicht immer möglich sein: sei es, weil wir uns des Vorhandenseins weiterer Dimensionen einfach nicht bewußt sind, sei es, weil es praktisch einfach nicht geht. Daß es "nicht geht", damit meinen wir eine gerade in der Marktforschung häufig auftretende Situation. Theoretisch wäre es z.B. durchaus möglich, von einer Seifensorte eine Reihe von Proben zu liefern, die sich entweder nur durch die Stärke des Geruchs, die Farbe usw.

unterscheiden. Praktisch wird jedoch meist von der Produktion her ein Satz von Proben vorgegeben, die die verschiedensten Kombinationen der verschiedensten Ingredienzien enthalten. Man soll der Marktforscher herausfinden, welche "am besten geht" und welche der Komponenten dafür verantwortlich sind. In einem solchen Fall müßte eine mehrdimensionale Skalierung durchgeführt werden.¹⁾

Wie wäre eine solche mehrdimensionale Skalierung möglich, wie könnte man das vorliegende Modell für den Fall mehrdimensionaler Kontinua erweitern?

b. Den besten Ansatzpunkt für eine derartige Erweiterung dürften die komparativen Methoden der topologischen Skalierung bieten. Interessante Untersuchungen von Coombs und seinen Schülern²⁾ sind zu diesem Thema im Entstehen. Die Ergebnisse dieser Untersuchungen - wenn sie erst einmal vorliegen - könnten wir u.U. für unser Modell adaptieren. Das Problem ist bis jetzt allerdings noch weit davon entfernt, gelöst zu werden. Wir wollen hier nur kurz auf den Weg hinweisen, auf dem man zu einem derartigen mehrdimensionalen Modell gelangen könnte.

Im eindimensionalen Fall hatten wir angenommen, daß sowohl Befragte wie Stimuli als Punkte auf einem eindimensionalen

1) Auch eine mehrdimensionale Faktoranalyse wäre unter gewissen Voraussetzungen möglich.

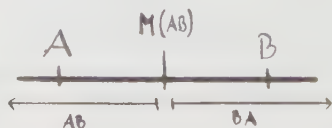
2) Coombs, G.H.: Psychological Measurement and A Theory of Data, 1958, Manuskript (tentative draft).
Bennett, J.F. and W.L. Hays: Multidimensional Unfolding: Determining the Dimensionality of Ranked Preference Data, Manuskript.
Hays, W.L. and J.F. Bennett: Multidimensional Unfolding: The Problem of Configuration, Manuskript.

Kontinuum, also einer Geraden, dargestellt werden können. Jeder Befragte wird die Stimuli entsprechend seiner eigenen Position anordnen - diese Anordnung bildet seine individuelle Präferenzskala. Die Aufgabe bestand nun darin, aus den empirisch gegebenen individuellen Präferenzskalen die Positionen der Befragten auf dem Kontinuum zu rekonstruieren.

Dieser Grundgedanke der Methode läßt sich sehr leicht auf den mehrdimensionalen Fall übertragen. Statt sich Stimuli und Befragte als Punkt auf einer Geraden zu denken, stellen wir sie uns jetzt als Punkte in einem mehrdimensionalen Raum vor. Wieder werden die individuellen Präferenzskalen der Befragten durch die relativen Entfernungen ihrer Position von der Position der Stimuli bestimmt. Jetzt ergibt sich aber die Aufgabe, aus diesen individuellen Präferenzskalen sowohl die Dimensionalität des Raums, wie die Projektion der Positionen der Stimuli und Befragten auf die Achsen dieses Raums zu bestimmen.

Zwei Stimuli A und B kann ein Befragter zum Beispiel AB, BA oder $A = B$ anordnen. Die Präferenzskala $A = B$ wird nur derjenige abgeben, von dessen Position aus der Abstand zu A gleich dem zu B ist. Im eindimensionalen Fall ist der Ort dieses Befragten der Mittelpunkt der Strecke $M(A,B)$. Alle Befragten, die sich rechts von diesem Punkt befinden, werden BA melden, alle, die sich links davon befinden, AB.

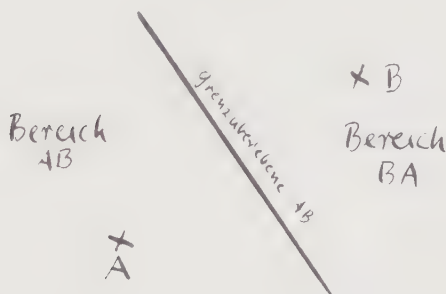
Schaubild 24



Bei mehreren Stimuli auf einem eindimensionalen Kontinuum ist nun das ganze Kontinuum durch eine Reihe solcher Mittelpunkte in Abschnitte geteilt, innerhalb derer alle Befragten eine bestimmte Präferenzskala besitzen werden.

Im zweidimensionalen Fall nun ist der Ort gleicher Abstände zwischen zwei Punkten das Lot auf dem Mittelpunkt der Verbindungsgeraden dieser zwei Punkte.

Schaubild 25



Alle Befragten, die sich auf der einen Seite des Lots befinden, werden AB, die anderen BA angeben. Diejenigen, deren Position genau auf dem Lot ist, werden $A = B$ melden.

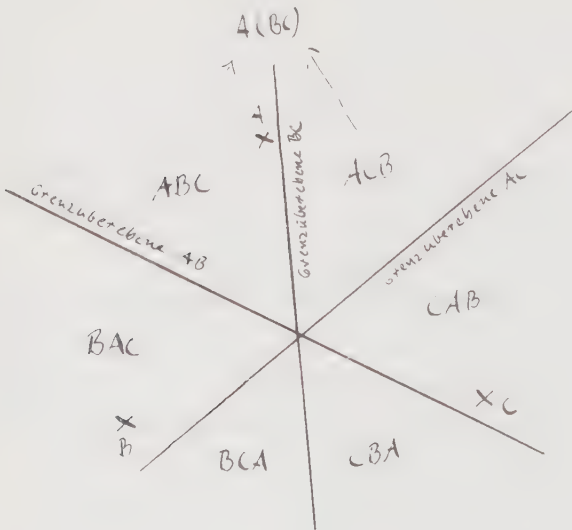
Im Fall eines r -dimensionalen Kontinuums wird nun dieser Ort gleicher Abstände eine Überebene von der Dimension $r-1$ sein. Jede dieser "Grenz-Überebenen" teilt den Raum in zwei "Halbräume", die Bereiche AB und BA, innerhalb derer wieder die Präferenzen AB bzw. BA abgegeben werden.

Geht nun eine weitere Grenz-Überebene $G(BC)$ durch den Bereich AB, so wird nun jeder Befragte, dessen Position in dem in dem Bereich AB zugleich auf der B Seite von $G(BC)$ liegt, die Präferenz ABC abgeben usw.

Auf diese Art und Weise lassen sich eine große Zahl von Unterbereichen bilden, die von Segmenten der übergebenen begrenzt werden und jeweils eine Gruppe von Befragten mit ganz bestimmten Präferenzskalen umfassen.

Dabei wird nun jeder dieser Unterbereiche durch eine Kombination sich überschneidender "Elementarbereiche" gebildet. "Elementarbereiche" bezeichnet dabei alle jene Bereiche, die durch eine einfache Anordnung von Stimuli charakterisiert werden. In Schaubild 26 wird zum Beispiel der Bereich A(BC) aus den Elementarbereichen ABC und ACB, C(AB) aus den Elementarbereichen CBA und CAB gebildet. Jeder Unterbereich der Elementar-

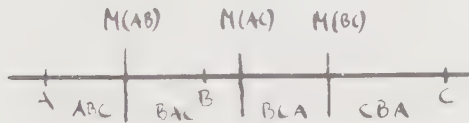
Schauold 26



plus einer zusätzlichen Verfeinerung der Anordnungen. Natürlich können sich Elementarbereiche nicht schneiden.

Wenn aber nun in einem r -dimensionalen Raum $r+2$ oder mehr Stimuli gegeben sind, werden einige der $(r+2)!$ "möglichen" Elementarbereiche verschwinden. Bei drei Stimuli in einem eindimensionalen Raum zum Beispiel (Schaubild 27) sind schon die "möglichen" Bereiche ACB und CAB nicht mehr vorhanden.

Schaubild 27



Die Abwesenheit dieser Elementarbereiche wird aber bestimmt durch die Dimensionalität des Raumes und die Position der Befragten in diesem Raum. Das heißt aber, daß wir umgekehrt diese Abhängigkeit als Mittel zur Rekonstruktion der Dimensionalität und der Position der Befragten benutzen können.

Soweit der Hinweis auf den Weg, der zu einer mehrdimensionalen Skalierung führen kann. Ob er gangbar ist, sollte Gegenstand weiterer Forschungen sein.

.

Literaturverzeichnis
=====

1. Anderson, O.: Probleme der statistischen Methoden-
lehre in den Sozialwissenschaften,
3.Aufl., Würzburg, 1957.
2. Bergmann, G. and K.W. Spence: The logic of psychological
measurement, Psych.Rev. (51), 1944,
S.1-24.
3. Bennett, J.F. and W.L. Hays: Multidimensional Unfolding:
Determining the Dimensionality of Ran-
ked Preference Data, vervielfältigtes
Manuskript, o.J. (ca. 1958/59).
4. Blankenship, A.B.: How to Conduct Consumer and Opinion
Research. The Sampling Survey in
Operation, New York, London, 1946.
5. Bradford, E.S.: Marketing Research: How to Analyse
Products, Markets and Methods of
Distribution, New York, 1951.
6. Campbell, N.: An Account of the Principles of
Measurement and Calculation, London -
New York, 1928.
7. Carnap, R.: Physikalische Begriffsbildung, Karls-
ruhe, 1926.
8. Comrey, A.L.: An operational approach to some
problems in psychological measurement
Psych.Rev. (57), 1950, S.217-228.

9. Coombs, C.H.: Psychological Measurement and A Theory of Data, Manuskript (tentative draft) May 1958.
10. Coombs, C.H.: The Theory and Methods of Social Measurement, Manuskript, o.J.
11. Coombs, C.H.: Application of the Unfolding Technique to the Study of Stimulus Scales, Manuskript, 1953.
12. Coombs, C.H.: The Measurement of Social Utility, Manuskript, 1953.
13. Deming, W.E.: Some Theory of Sampling, New York, 1950.
14. Deming, W.E.: Sample Decision in Business Research, New York, 1960.
15. Edwards, A.L.: Techniques of Attitude Scale Construction, New York, 1957.
16. Edwards, A.L. and F.P. Kilpatrick: A technique for the construction of attitude scales, Journ. appl. Psych. (32), 1948, S.374-384.
17. Edwards, A.L. and K.C. Kenney: A comparison of the Thurstone and Likert techniques of attitude scale construction, Journ. appl. Psych. (30), 1946, S.72-83.
18. Fechner, C.T.: Elemente der Psychophysik, Leipzig, 1860.
19. Guilford, J.P.: The phi coefficient and chi square as indices of item validity, Psychometrika (6), 1941, S.11-19.

20. Guttman, L.: The quantification of a class of attributes: A theory and method of scale construction. In Horst, Paul et al: The Prediction of Personal Adjustment, New York, 1941.
21. Guttman, L.: Kapitel 2,3,6,8 und 9 in Stouffer et al: Measurement and Prediction, Princeton, 1950.
22. Guttman, L.: The principal components of scalable attitudes, in P.F. Lazarsfeld: Mathematical Thinking in the Social Sciences, Glencoe, Ill., 1954.
23. Hobart, D.M.: Praxis der Marktforschung, Essen, 1952.
24. Hays, W.L. and J.F. Bennett: multidimensional Unfolding: The Problem of Configuration. Manuskript, o.J. (ca. 1958/59).
25. Hays, W.L.: A Proposal for the Non-metric Analysis of Multidimensional Similarities Data, Manuskript, o.J.
26. Kellerer, H.: Theorie und Technik des Stichprobenverfahrens, München, 1953.
27. Kendall, M.C.: Rank Correlation Methods, London, 1929 (2.Aufl.).
28. Kendall, M.C.: The Advanced Theory of Statistics, London, 1952.
29. Koller, S.: Zur Problematik des statistischen Messens, Allg.Stat.Arch.(40), 1956, S.316-340.

30. Lazarsfeld, P.F. (ed.): Mathematical Thinking in the Social Sciences, Glencoe, Ill., 1954.
31. Likert, R.: ^{measurement of} A technique for the/attitudes, Arch. Psych. (140), 1932.
32. Loevinger, J.: The technic of homogeneous tests compared with some aspects of "scale analysis" and factor analysis, Psych. Bull. (45), 1948, S.507-530.
33. Mosteller, F.: Remarks on the method of paired comparison, I, II, III, Psychometrika (16), 1951, S.3-11, 203-206, 207-218.
34. Neyman, J.: On two different aspects of the representative method: the method of stratified sampling and the method of purposive selection, J.R.S., 1934.
35. Pfanzagl, J.: Die axiomatischen Grundlagen einer allgemeinen Theorie des Messens, Würzburg, 1959.
36. Riley, M. et al.: Sociological Studies in Scale Analysis, New Brunswick, N.J., 1954.
37. Stevens, S.S.: Mathematics, measurement and psychophysics, in S.S. Stevens (ed): Handbook of Experimental Psychology, New York, 1951.
38. Stevens, S.S.: Measurement, psychophysics and Utility in Churchman and Ratoosh (ed.): Measurement: Definitions and Theories, New York, 1959.

39. Suchman, E.A.: Kapitel 4,5 und 7 in Stouffer et.al.:
Measurement and Prediction, Princeton,
1950.
40. Tagliacarne, G.: Tecnica e pratica delle ricerche di
mercato, Mailand, 1951.
41. Thurstone, L.L.: A law of comparative judgment,
Psych.Rev. (34), 1927, S.273-286.
42. Thurstone, L.L.: Psychophysical Methods in T.C.Andrews
(ed.): Methods of Psychology, New York,
1948.
43. Thurstone, L.L.: The measurement of values, Psych.Rev.
(1), 1951, S.47-58.
44. Thurstone, L.L. and E.J. Chave: The Measurement of Attitude,
Chicago, 1929.
45. Torgerson, W.S.: Theory and Methods of Scaling, New
York, 1958.
46. Winkler, W.: Statistische Maße, die ungleich messen,
Allg. Stat. Arch. (40), 1956, S.297-315.

* * * * *

Lebenslauf

Ich wurde am 15. Oktober 1934 als Sohn des Verwaltungsbeamten Paul Wehowski und seiner Ehefrau Helene in Breslau geboren. Ich bin deutscher Staatsangehöriger.

Am 14. Juli 1953 bestand ich die Reifeprüfung der Oberrealschule Deggendorf. Anschließend studierte ich Rechtswissenschaft, Betriebswirtschaft und Volkswirtschaft an den Universitäten München (9 Semester) und Münster (2 Semester) sowie an der Hochschule für Welthandel in Wien (1 Semester). Am 8. November 1958 bestand ich das Diplomexamen für Volkswirte an der Universität München.

Lebenslauf

Ich wurde am 15. Oktober 1934 als Sohn des Verwaltungsbeamten Paul Wehowski und seiner Ehefrau Helene in Breslau geboren. Ich bin deutscher Staatsangehöriger.

Am 14. Juli 1953 bestand ich die Reifeprüfung der Oberrealschule Deggendorf. Anschließend studierte ich Rechtswissenschaft, Betriebswirtschaft und Volkswirtschaft an den Universitäten München (9 Semester) und Münster (2 Semester) sowie an der Hochschule für Welthandel in Wien (1 Semester). Am 8. November 1958 bestand ich das Diplomexamen für Volkswirte an der Universität München.

